

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS



**METODOLOGÍA BAYESANA APLICADA AL
ESTUDIO DE CONFIGURACIONES ESPACIALES
EN BIOINFORMÁTICA.**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR**

Carmen Nieto Zayas

Bajo la dirección del doctor

Juan Miguel Marín Diazaraque

Madrid, 2010

• **ISBN: 978-84-693-4093-6**

© Carmen Nieto Zayas, 2009

METODOLOGÍA BAYESIANA
APLICADA AL ESTUDIO
DE CONFIGURACIONES ESPACIALES
EN BIOINFORMÁTICA

Tesis Doctoral
Carmen Nieto Zayas

Dirigida por:
Juan Miguel Marín Diazaraque

Mayo 2009

Juan Miguel Marín Diazaraque, Profesor del Departamento de Estadística de la Universidad Carlos III de Getafe, Madrid,

CERTIFICA:

Que la presente memoria titulada

**METODOLOGÍA BAYESIANA APLICADA AL ESTUDIO DE
CONFIGURACIONES ESPACIALES EN BIOINFORMÁTICA**

ha sido realizada bajo mi dirección por Dña. Carmen Nieto Zayas, licenciada en C.C. Matemáticas, y constituye su Tesis para optar al Grado de Doctor en Ciencias Matemáticas.

Y para que coste, en cumplimiento de la legislación vigente y a los efectos oportunos, firmo la presente en Madrid, a 28 de Mayo de 2009.

Fdo: J. Miguel Marín Diazaraque

Agradecimientos

Principalmente quiero dar las gracias a una persona que conozco desde hace más de quince años. El azar hizo que fuera mi compañero de despacho en la Universidad durante varios años lo que permitió que nos convirtiéramos en amigos. A pesar de que cambió de destino, siempre mantuvimos el contacto hasta que un día me propuso dirigirme la tesis. Muchas gracias Miguel, porque además de ser un gran amigo y una gran persona, tuviste la osadía de ser mi director de tesis y, gracias a ti, he conseguido finalizar este trabajo que es tanto tuyo como mío.

Tengo la gran suerte de tener buenos amigos de los que he recibido un gran apoyo y comprensión a lo largo de todo este tiempo. Muchas gracias amiga Inés, amiga Julia, amiga María, amiga M^a Jesús, amigo Pedro, amiga Pepa y amiga Tere, porque nadie mejor que vosotros entendéis lo que es emprender este largo y, en ocasiones, duro camino de la tesis.

Gracias, Carlos, que has sabido pacientemente adaptarte a mi dedicación a esta tesis, a mi disciplina en el estudio y hacer de padre y madre en momentos donde la tesis me reclamaba. Por último, quiero agradecer a mi hijo Alonso que, aun a su corta edad y a su manera, se preocupaba de mi tesis y con una sonrisa me preguntaba "*¿te falta mucho?*". Espero haber sido un ejemplo de que con el esfuerzo todo se consigue.

Resumen

En esta tesis se estudia la superposición y el emparejamientos de configuraciones o conjuntos de puntos en el espacio, considerando distintas transformaciones geométricas entre las mismas. El estudio de imágenes médicas o la comparación estructural de pares de moléculas son algunos ejemplos de problemas de configuraciones de puntos. En una primera parte se extiende el modelo bayesiano de Green y Mardia (2006) en dos sentidos: por un lado, se consideran los emparejamientos entre más de dos configuraciones y, por otro, se asumen transformaciones geométricas entre ellas más generales. En una segunda parte, se estudian transformaciones no lineales entre configuraciones, planteando dos modelos de redes neuronales con los que se analiza, desde la perspectiva bayesiana, la relación entre dos configuraciones etiquetadas.

Un aspecto importante de esta tesis es su aplicación en el campo de la Bioinformática. Se parte de datos procedentes de microarrays, consistentes en las expresiones de los genes que se quieren estudiar. Una aportación novedosa es el tratamiento de estas expresiones hasta llegar a una representación de los genes como configuraciones de puntos en un espacio. Esto se lleva a cabo mediante la técnica multivariante multidimensional scaling ponderada (*INDSCAL*), utilizada en el ámbito de las ciencias sociales, pero menos frecuente en el contexto de la Bioinformática, permitiendo un estudio geométrico de los genes a través de su disposición en el espacio.

Índice general

Prólogo	VII
1. Motivación y Antecedentes	1
1.1. Introducción al análisis de datos espaciales	1
1.2. El análisis de la forma, superposición y emparejamiento, como problemas espaciales	3
1.2.1. Métodos tradicionales frente a métodos geométricos	4
1.3. Los objetos como configuraciones de puntos en el espacio	6
1.4. La superposición de configuraciones: Métodos Procrustes	11
1.5. El estudio de configuraciones no etiquetadas	11
2. Definiciones y conceptos previos	17
2.1. Introducción	17
2.2. Proceso de Poisson espacial	18
2.2.1. Introducción	18

2.2.2.	Los procesos puntuales espaciales	19
2.2.3.	El Proceso de Poisson espacial	20
2.3.	Transformaciones matriciales	26
2.3.1.	Transformaciones matriciales simples	26
2.3.2.	Composición de transformaciones matriciales simples	34
2.4.	El emparejamiento de dos configuraciones de puntos no etiquetadas .	37
2.4.1.	El modelo matemático	38
2.4.2.	El Proceso de Poisson de la configuración de referencia	39
2.4.3.	La matriz de emparejamientos y su distribución a priori	39
2.4.4.	Verosimilitud de los datos	40
2.4.5.	Inferencias sobre los parámetros	41
2.5.	Introducción a los modelos estadísticos de redes neuronales artificiales	43
2.5.1.	El modelo biológico	43
2.5.2.	Estructura de una red neuronal artificial	44
2.5.3.	Los modelos de regresión como redes neuronales	45
2.5.4.	Redes neuronales Perceptrón.	46
2.6.	Escalamiento multidimensional: análisis INDSCAL	51
3.	Emparejamiento de más de dos configuraciones no etiquetadas mediante transformaciones lineales	53

3.1. Introducción	53
3.2. Modelización del problema	54
3.3. El Proceso de Poisson espacial de la configuración de referencia . . .	56
3.4. Matrices de afinidad M y matrices de emparejamiento S	58
3.4.1. Matriz de afinidad y matriz de emparejamientos de orden 2 . .	58
3.4.2. Matriz de afinidad y matriz de emparejamientos de orden 3 . .	59
3.4.3. Matriz de afinidad y matriz de emparejamientos de orden k . .	60
3.4.4. Distribuciones a priori de las matrices de afinidad y de em- parejamientos	63
3.5. Modelización con transformaciones lineales generales	68
3.5.1. Construcción de la verosimilitud de los datos	69
3.5.2. Emparejamiento de $M = 4$ configuraciones con errores normales	73
3.5.3. Inferencia bayesiana en el caso de emparejamientos conocidos de $M = 4$ configuraciones	79
3.6. Modelización con transformaciones euclídeas de similitud	84
3.6.1. Construcción de la verosimilitud de los datos	86
3.6.2. Emparejamientos de $M = 4$ configuraciones con errores nor- males	86
3.6.3. Inferencia bayesiana en el caso de emparejamientos conocidos de $M = 4$ configuraciones	88

3.6.4.	Inferencia bayesiana en el caso de emparejamientos desconocidos de $M = 3$ configuraciones y transformaciones rígidas . . .	99
3.7.	Aplicaciones	107
3.7.1.	Datos simulados	108
3.7.2.	Una aplicación en Bioinformática	110
4.	Emparejamiento de pares de configuraciones mediante transformaciones no lineales	125
4.1.	Introducción	125
4.2.	Emparejamiento de configuraciones no etiquetadas bajo modelo de redes neuronales basado en un Proceso de Poisson	126
4.2.1.	Modelización del problema	126
4.2.2.	Construcción de la verosimilitud de los datos	129
4.2.3.	Inferencia bayesiana en el caso de configuraciones etiquetadas	136
4.3.	El modelo con $M = 1$ nodo, en el caso de \mathbb{R}^2	147
4.3.1.	Inferencia bayesiana en el caso de configuraciones etiquetadas	150
4.4.	Emparejamiento de configuraciones etiquetadas bajo modelo de redes neuronales con una configuración fija	154
4.4.1.	Modelización del problema	155
4.4.2.	Construcción de la verosimilitud de los datos	156
4.4.3.	Expresiones de las distribuciones a posteriori de los parámetros	157

4.5. Una aplicación en Bioinformática	162
4.5.1. Representación de los genes	163
4.5.2. Elección del número de nodos del modelo	164
4.5.3. Aplicación del modelo de redes neuronales basado en un Pro- ceso de Poisson	165
4.5.4. Aplicación del modelo de redes neuronales con una configura- ción de referencia	171
Futuras líneas de investigación	179
Referencias	181

Prólogo

El trabajo realizado en esta tesis aborda dos líneas de investigación centradas en el estudio de configuraciones o conjuntos de puntos en el espacio. Por un lado, extiende el modelo bayesiano jerárquico de Green y Mardia (2006) sobre el emparejamiento de dos configuraciones no etiquetadas de puntos, al caso de más de dos configuraciones, además de asumir transformaciones geométricas más generales entre ellas. En segundo lugar, se estudian transformaciones no lineales entre configuraciones, planteando dos modelos de redes neuronales que permiten estudiar la relación entre dos configuraciones etiquetadas, utilizando la metodología bayesiana.

La asignación de etiquetas a los puntos de las configuraciones para su identificación, puede hacerse de forma aleatoria en cada configuración, de manera que no existe ninguna correspondencia entre puntos de distintas configuraciones, o bien, se puede utilizar algún criterio, de forma que haya puntos de distintas configuraciones a los que, según este criterio, se les asigne las mismas etiquetas, quedando así emparejados. En el primer caso se dice que las configuraciones son *no etiquetadas* y en el segundo que sí lo son.

Con respecto a la primera parte de esta tesis, la modelización de M configuraciones no etiquetadas se plantea bajo la suposición de que las configuraciones son perturbaciones aleatorias de un conjunto de puntos fijos y desconocidos $\{\mu_l\} \in \mathbb{R}^d$, la configuración de referencia, proveniente de un proceso de Poisson, y que cada μ_l sólo puede generar a lo sumo un punto de una misma configuración pero más de un punto de distintas configuraciones. Se desconoce qué puntos de cada configuración se

corresponden con cada uno de estos μ_l , pero aquellos puntos de configuraciones distintas que son perturbaciones aleatorias de un mismo punto μ_l estarán emparejados. Por otra parte, se asume que existen unas transformaciones geométricas desconocidas entre los sistemas de referencia donde se representan las M configuraciones, que también formarán parte del modelo.

Se definen unas *matrices de emparejamientos y de afinidad* de distintos órdenes de emparejamientos, que modelizan los emparejamientos entre configuraciones. Se plantean los modelos con transformaciones lineales generales y con transformaciones de similitud entre las configuraciones, particularizando al caso de $M = 4$ configuraciones con distribución normal de los errores, para realizar inferencias sobre los parámetros, en el caso de conocer los emparejamientos. En el caso de transformaciones de similitud se realizan inferencias cuando son configuraciones no etiquetadas.

Sobre la segunda parte de esta tesis, continuando con la misma modelización de configuraciones no etiquetadas, se aborda el problema de asumir transformaciones no lineales de redes neuronales entre pares de configuraciones, aunque las inferencias sobre los parámetros se desarrollan sólo en el caso de emparejamientos conocidos. También se considera un segundo modelo considerando que una de la dos configuraciones es fija y la otra una transformación no lineal de la primera más un error aleatorio.

Un aspecto importante de esta tesis es su aplicación y el ámbito en el que se lleva a cabo. El análisis de la forma de objetos, el estudio de imágenes médicas, la comparación estructural de pares de moléculas son algunos ejemplos de problemas de configuraciones de puntos. Las aplicaciones presentadas en esta tesis se centran en el ámbito de la Bioinformática, donde la información de la que se parte son datos de microarrays. En particular se utilizan dos bases de datos. Una primera tomada de un experimento de Karaman et al. (2003) con células cultivadas de fibroblastos en tres especies genéticamente relacionadas: gorilas (*Gorilla gorilla*), bonobos (*Pan paniscus*) y humanos (*Homo sapiens*). La segunda, es la base de datos de microarrays ALL (Acute Lymphoblastic Leukemia) de Chiaretti et al. (2004) de pacientes con

leucemia linfoblástica aguda.

El tratamiento de estas expresiones hasta llegar a una representación de los genes en un espacio, es una de las aportaciones novedosas de este trabajo, utilizando la técnica multivariante multidimensional scaling ponderada (*INDSCAL*) cuya aplicación, aunque utilizada en el ámbito de las ciencias sociales, en Bioinformática es innovadora.

La aplicación de los modelos estudiados a configuraciones de genes permite, haciendo una selección adecuada de los mismos, realizar un estudio de ellos desde un punto de vista geométrico. Se podría, por ejemplo, estudiar la evolución de una enfermedad comparando la disposición de genes relevantes en diferentes instantes del transcurso de la misma, o comparar la expresión de los genes en pacientes sanos y pacientes enfermos.

Capítulo 1

Motivación y Antecedentes

1.1. Introducción al análisis de datos espaciales

La investigación con datos recogidos en el espacio está creciendo en disciplinas tan distintas como la ecología, marketing, medicina, climatología, bioinformática... Son muy diversos los tipos de investigación que es posible realizar y dependen, en gran medida, del tipo de datos espaciales de los que se disponga. Estos datos pueden corresponder a vectores aleatorios observados en distintas localizaciones y a las coordenadas de estas posiciones. Además es posible que se recoja la información en distintos instantes de tiempo, de manera que se trate de investigaciones espacio-temporales.

En función de la información disponible y del carácter de la región en el espacio donde se centre el estudio, los conjuntos de datos espaciales se pueden clasificar en tres tipos:

1. *Datos geoestadísticos*: Fijada una región $D \subset \mathbb{R}^d$ que contiene un rectángulo d -dimensional de volumen positivo, los datos corresponden a las observaciones de un vector aleatorio $Y(\mathbf{s})$ en cualquier localización $\mathbf{s} \in D$.

Tal podría ser el caso en el que se recogen las concentraciones de un sustrato

en una región de terreno.

2. *Datos reticulares*: La región $D \subset \mathbb{R}^d$ es una colección numerable de puntos de \mathbb{R}^d , regular o irregular, en cada uno de los cuales se recogen las observaciones del vector aleatorio $Y(\mathbf{s})$. La región suele construirse teniendo forma de retícula, aunque en ocasiones presenta formas irregulares. Estas localizaciones a menudo representan los centroides de zonas en las que está dividida la región D y, conectadas entre sí, adquieren un aspecto de retícula. La observación $Y(\mathbf{s})$ recoge la información de toda el área de estudio representada por su centroide \mathbf{s} .

Dos situaciones de datos reticulares se encuentran en las imágenes proporcionadas por los satélites y en las imágenes médicas que, aunque las escalas espaciales en ambos problemas son muy diferentes, la forma de los datos y las cuestiones en ambos contextos son similares.

3. *Datos de un proceso puntual*: A diferencia de los casos anteriores, la región D es aleatoria, de manera que los datos corresponden a las localizaciones de los eventos aleatorios encontrados en D . El valor de $Y(s)$ es fijo, y representa la ocurrencia o no del evento de interés. También es posible recoger, además de la localización donde ha tenido lugar el evento, información adicional de alguna variable, obteniendo lo que se denomina un proceso puntual con marcas.

La localización de personas con una enfermedad determinada en cierta zona, o la de ciertas especies en Biología, son ejemplos de este tipo de datos.

Aunque el objetivo de las investigaciones es muy diferente, un contexto especial donde también se trabaja con datos espaciales es el análisis de formas de objetos. Las localizaciones de los puntos vienen determinadas por la forma del objeto en cuestión, ya que se sitúan en puntos estratégicos (*landmarks*) del contorno del objeto que permiten describir la forma que tiene. Esto permite, por ejemplo, estimar la forma media de una población de objetos a partir de los landmarks de una muestra aleatoria de ellos, o comparar las formas medias de objetos distintos. No obstante, el estudio

de la forma se ha extendido a conjuntos de puntos aleatorios, como proyecciones de partículas de polvo o de granos de arena (Stoyan y Molchanov (1997)).

Pero además de la comparación de la forma de dos o más conjuntos de puntos en el espacio, es interesante plantearse la búsqueda de transformaciones que mejor los superpongan y, en el caso de que no se conozca la correspondencia entre los puntos, investigar métodos de estimación de emparejamientos.

1.2. El análisis de la forma, superposición y emparejamiento, como problemas espaciales

El análisis, descripción y comparación de formas y tamaños de los objetos y seres que nos rodean, han sido y son actualmente de gran interés en muchas disciplinas. Se denomina *morfometría* el estudio estadístico de la forma y el tamaño de un objeto y sus relaciones con otras variables y es una herramienta muy utilizada en Arqueología, Paleontología, Geografía, Geología, o en investigaciones forenses. Así, por ejemplo, se estudia de qué manera cambia la forma del cuerpo humano a lo largo del tiempo, o si es posible saber por la forma o el tamaño de un cráneo, si éste perteneció a un hombre o a una mujer.

En Medicina y Biología tiene especial interés estudiar cómo la forma de un organismo cambia durante el crecimiento o a lo largo de la evolución, cómo viene relacionada con el tamaño o con otras variables como el sexo, la edad o condiciones ambientales en las que se haya desarrollado el organismo. Ya Galileo, en 1638, observó que los huesos de animales grandes no son exactamente versiones a mayor escala de esos mismos huesos en animales más pequeños. La explicación viene dada porque los animales de mayor peso necesitan huesos proporcionalmente más gruesos para soportar el incremento de peso. Otra aplicación muy interesante es poder discriminar y clasificar organismos utilizando la forma que estos presentan.

El análisis estadístico de formas recoge la metodología para analizar las formas

en condiciones de aleatoriedad. Los objetos a estudiar podrían ser considerados como una muestra aleatoria de una población, siendo los objetivos principales estimar la forma media de la población, estimar la variabilidad de las formas e incluso realizar inferencias como, por ejemplo, comparar las formas medias de dos poblaciones de objetos.

1.2.1. Métodos tradicionales frente a métodos geométricos

Los métodos que se han utilizado para el análisis de la forma de objetos han ido cambiando a lo largo del tiempo. En los últimos años se ha desarrollado toda una teoría del análisis de las formas desde un punto de vista geométrico, centrándose los estudios en la definición de la forma a través de ciertos puntos claves del objeto que lleguen a caracterizarlo no perdiendo la estructura geométrica del mismo. Este enfoque tiene su origen en el trabajo de Thompson (1917) que propuso representar la forma de los objetos en una rejilla con el objetivo de, mediante deformaciones de la misma, describir cambios morfológicos en ellos.

A principios del siglo XX se abordó este problema, desde otro punto de vista, basándose en la descripción de los objetos a partir de mediciones cuantitativas de los mismos (longitud, altura, anchura, distancia entre ejes, cociente de longitudes...), siendo desarrolladas numerosas técnicas multivariantes hoy en día ampliamente conocidas y utilizadas. Este enfoque es conocido en biología como *morfometría multivariante*. Pearson, en 1926, estudió un coeficiente de semejanza racial como medida de similitud entre cráneos basada en las distancias de un gran número de landmarks. Este término ya había sido definido en 1882, en el décimotercer congreso de la Sociedad Antropológica Alemana.

Algunas aplicaciones directas de este enfoque en el análisis de formas, son la clasificación de especies e identificación del sexo basándose en mediciones cuantitativas craneales. El análisis de componentes principales sobre las mediciones es una técnica que ampliamente se ha utilizado en este contexto, y cada componente se

identificaba con un aspecto del tamaño o la forma. Frecuentemente la primera componente principal se interpretaba como una medida global del tamaño del objeto, pues ponderaba con el mismo peso a todas las variables cuantitativas consideradas.

Otra área de gran aplicación de estas técnicas multivariantes es la *alometría*, que consiste en el estudio de las diferencias en formas asociadas al tamaño. El concepto de alometría fue desarrollado por Huxley en 1924.

Mosimann, en 1970, realizó un análisis en profundidad en esta materia, aportando teoremas donde se estudia la independencia entre el tamaño y la forma, concepto que denomina *isometría*. Como ejemplo, estudia la relación entre el tamaño y la forma en el hombre. Define la variable adimensional $\frac{H}{B}$ con la que se pretende recoger la idea de forma, donde H es la longitud de la cabeza (de la barbilla hasta la parte más alta) y B la longitud del cuerpo, excluyendo la cabeza. Por otra parte, define la variable tamaño $H + B$, dimensional, y que representa la altura del individuo. Demuestra, a partir de una muestra de personas, tanto adultos como niños, que la correlación entre la variable de forma $\frac{H}{B}$ y la variable de tamaño $H + B$ es negativa.

Toda esta teoría está basada exclusivamente en variables positivas (longitudes, ángulos, cociente de longitudes...). Hay que señalar que la idea de considerar las coordenadas de las localizaciones en vez de las distancias, ángulos, etc, entre ellas, permite disponer de mayor información además de poder recurrir, si fuera necesario, a cualquier medida cuantitativa que se necesite.

En muchas ocasiones, la interpretación de las combinaciones lineales construidas a partir de las variables cuantitativas es poco intuitiva y complicada. Por todo esto, es interesante abordar el problema desde un punto de vista geométrico, partiendo del espacio original desde el que se obtienen las mediciones cuantitativas, es decir, trabajar en el espacio geométrico original del objeto en vez de en el espacio multivariante de sus mediciones.

Los nuevos avances tecnológicos han permitido desarrollar este nuevo enfoque espacial del problema gracias a la digitalización de objetos.

Estos problemas requieren dar una definición exhaustiva del concepto de forma, así como decidir qué aspectos geométricos de los objetos son prescindibles a la hora de definirla. Esto lleva a estudiar las transformaciones óptimas que superpongan a los objetos, ya que la localización, orientación o tamaño de un objeto no afecta a su forma. Los métodos *Procrustes* son ampliamente utilizados en este contexto y se basan en la estimación de transformaciones de similitud mediante mínimos cuadrados.

Por último, también un aspecto muy interesante en este contexto de datos espaciales, es la estimación de los emparejamientos de objetos cuando no se conoce qué puntos de uno corresponden con los del resto, por lo que el problema consiste en identificar qué puntos están emparejados. Problemas de este tipo se plantean con frecuencia en Bioinformática y Quimiinformática, donde las estructuras moleculares no permiten conocer qué partes de las moléculas se corresponden. Trabajos realizados en este contexto se pueden encontrar en Dryden et al. (2007) y en Green y Mardia (2006).

Como conclusión, los problemas de análisis de formas, de superposición y de emparejamientos de objetos requieren una representación geométrica de los mismos, dando lugar a investigaciones y análisis con datos espaciales.

1.3. Los objetos como configuraciones de puntos en el espacio

La forma de un objeto está relacionada con la apariencia o con la información geométrica que tiene el mismo, pero es difícil en ocasiones describirla. En el lenguaje cotidiano la forma desconocida de un objeto a menudo se relaciona con una segunda forma de otro al que se parece. Así, por ejemplo, se dice que el mapa de Italia tiene forma de bota. La siguiente definición de forma es debida a Kendall (1977) y es la más referida.

Definición 1 *Se entiende por forma a toda la información geométrica que permanece en un objeto prescindiendo de cambios en la posición, en la escala y en giros o rotaciones.*

En otras palabras, la forma es invariante a transformaciones euclídeas de similitud de localización, escala y rotación. Esta definición recoge la idea intuitiva de que independiente de dónde se localice el objeto, qué tamaño tenga y si está rotado de una manera o de otra. Así, dos objetos tendrán la misma forma si después de ciertas transformaciones de traslación, de escala y de rotación en un sistema de coordenadas arbitrario, coinciden, es decir, son objetos similares.

En la Figura 1.1 las dos siluetas de la luna tienen la misma forma aunque distinta localización y tamaño. En la Figura 1.2 las dos siluetas tienen la misma forma y el mismo tamaño, pero distinta localización.

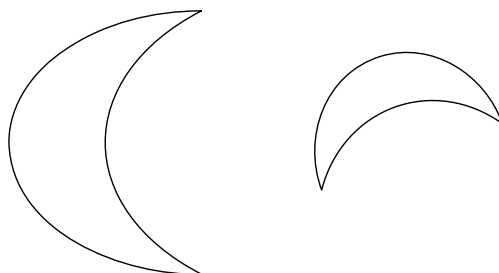


Figura 1.1 Dos siluetas de luna con la misma forma

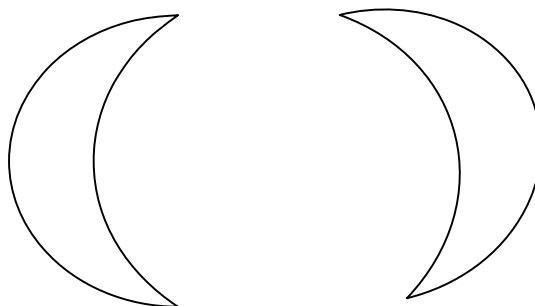


Figura 1.2 Dos siluetas de luna con la misma forma y tamaño

En la práctica, tiene interés comparar objetos con distintas formas, por lo que se necesita alguna manera de describirlas. Una manera de hacerlo es localizando

un número finito de puntos en su entorno, denominados *landmarks* (Dryden and Mardia, 1992).

Definición 2 *Un landmark es un punto de un objeto u organismo que caracteriza o es común a los de su misma población o especie.*

Los landmarks van a ayudar a describir la forma de un objeto localizando un número finito de ellos sobre él. Dryden and Mardia (1992) definen tres tipos de landmarks:

- i) *Landmarks anatómicos*: es un punto especial identificado por un experto que tiene algún significado específico dentro del objeto u organismo.
- ii) *Landmarks matemáticos*: se caracterizan por verificar alguna propiedad matemática o geométrica dentro del objeto (punto de máxima curvatura, punto máximo...)
- iii) *Pseudo-landmarks*: son puntos localizados normalmente entre dos landmarks anatómicos o matemáticos. Suelen dibujarse equidistantes y ayudan a aproximar curvas continuas del objeto a estudiar.

Definición 3 *Se llama configuración al conjunto de landmarks de un determinado objeto. Toda configuración se representa a través de la matriz de la configuración X , de dimensiones $k \times m$, que recoge las coordenadas cartesianas de k landmarks en dimensión m .*

Es frecuente asignar a cada landmark una etiqueta, que suele ser un nombre o un número, con el objetivo de identificarlo. Esto es importante en problemas de comparación y emparejamientos de objetos, pues ayudarán a describir qué pares de landmarks están emparejados.

La asignación de las etiquetas a los landmarks puede ser elegida según algún criterio, de manera que los landmarks de distintos objetos de la misma especie vendrán identificados con las mismas etiquetas. Se dice entonces que son landmarks etiquetados. Sin embargo, puede haber situaciones donde la asignación de las etiquetas a los landmarks sean arbitrarias y no exista una correspondencia de landmarks entre puntos de distintas especies. En este caso, se dice que los landmarks no están etiquetados. En esta situación, un interesante objetivo que se plantea es la estimación de los mejores emparejamientos de landmarks.

Definición 4 Sean $X = \{x_j, j = 1, \dots, n\}$ e $Y = \{y_k, k = 1, \dots, n\}$ dos configuraciones de n puntos en \mathbb{R}^d . Se dice que son dos configuraciones de puntos etiquetados, o son dos configuraciones etiquetadas, si se conoce qué puntos de la configuración X están emparejados con qué puntos de la configuración Y , es decir, se sabe cuáles son los emparejamientos entre sus puntos. En el caso de que se conozcan sólo los emparejamientos de algunos de los puntos, se dirá que son configuraciones parcialmente etiquetadas, o que tienen emparejamientos parcialmente conocidos.

Ejemplo 5 En O'Higgins (1989), se puede encontrar un ejemplo de configuraciones etiquetadas. Se tomó ocho landmarks anatómicos, seleccionados por un especialista, en la sección vertical media de los cráneos de simios para estudiar si existían diferencias entre los cráneos de simios hembra y simios macho. En las Figuras 1.3 y 1.4 se representan las configuraciones de dos de los simios hembras y de dos de los simios machos que utilizó en su estudio. Los nombres de las etiquetas corresponden a puntos identificativos de la sección vertical de los cráneos.

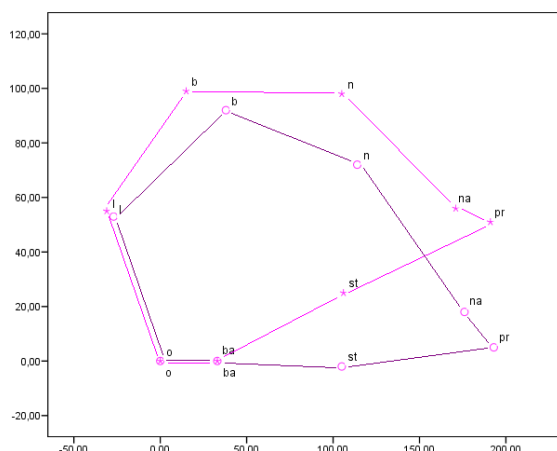


Figura 1.3 Dos configuraciones de cráneos de simios hembras

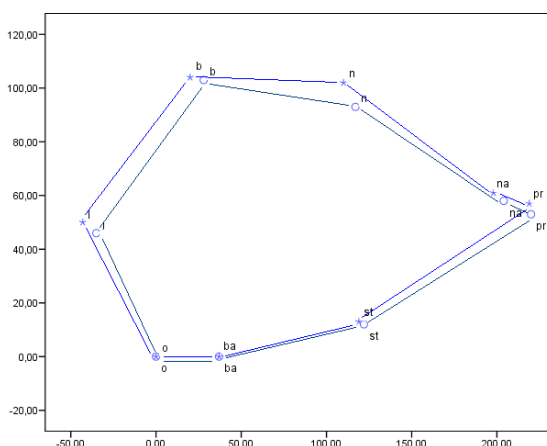


Figura 1.4 Dos configuraciones de cráneos de simios machos

El estudio de la superposición de dos o más configuraciones etiquetadas permite comparar sus formas, ya que éstas no se ven afectadas por sus localizaciones, orientaciones o tamaños. Así, se estudian las transformaciones que superponen a las configuraciones de puntos. Los métodos Procrustes son ampliamente utilizados en este contexto. En el caso de configuraciones no etiquetadas, el estudio se extiende también a la búsqueda de emparejamientos entre sus puntos además del análisis de la forma.

1.4. La superposición de configuraciones: Métodos Procrustes

El análisis Procrustes estudia la superposición de configuraciones etiquetadas mediante transformaciones de similitud (rotaciones, traslaciones y escalamientos), minimizando la distancia entre ellas mediante la técnica de mínimos cuadrados. No obstante, estos métodos se han ido adaptando al caso de configuraciones no etiquetadas o parcialmente etiquetadas, extendiendo la minimización al conjunto de las transformaciones y al conjunto de las permutaciones de los landmarks.

El origen del término se encuentra en la mitología griega. Procrustes fue el apodo de un ladrón que vivió en los caminos entre Atenas y Eleusis. Ofrecía a los viajeros una habitación para pasar la noche y los *ajustaba* a la cama estirándolos si eran demasiado bajos o acortándolos si eran demasiado altos. La analogía con respecto a las técnicas actuales puede verse interpretando una configuración como la cama que ofrecía Procrustes y la otra identificándola con la persona que iba a pasar la noche en ella, pues era *rotada*, *trasladada* y hasta *reescalada* para que se ajustara lo más posible a la cama. El término análisis Procrustes fue acuñado por primera vez por Hurley y Cattell en relación al Análisis Factorial (1962).

El análisis Procrustes utilizando matrices ortogonales (rotación-reflexión), inicialmente tuvo su aplicación en Psicología y las primeras publicaciones aparecieron en la revista *Psychometrika*.

1.5. El estudio de configuraciones no etiquetadas

En el caso de configuraciones no etiquetadas o parcialmente etiquetadas, además del estudio de su forma, es de gran interés buscar emparejamientos entre sus puntos y ajustar una transformación geométrica entre ellas.

Con respecto al análisis de la forma de una configuración no etiquetada, inicial-

mente, algunas investigaciones se centraron en el estudio de triángulos (configuraciones de tres puntos).

A principio de los años veinte, se planteó el estudio de la localización de 52 piedras megalíticas situadas en Cornwall, Inglaterra. El enfoque que se le dio al problema fue analizar las formas de todos los triángulos que se pudieran formar con las 52 localizaciones, en total $\binom{52}{3}$ triángulos, y comprobar si se obtenían más triángulos *planos*, es decir tres puntos casi alineados con el ángulo mayor cerca de 180 grados, que los que se esperaría encontrar bajo hipótesis de aleatoriedad. Los puntos en este problema no estaban etiquetados y los datos consistían en las posiciones de los $\binom{52}{3}$ triángulos en el plano. Estos datos fueron particularmente importantes porque motivaron investigaciones pioneras de D.G. Kendall. Trabajos relacionados con este problema se pueden encontrar en Broadbent (1980), Kendall y Kendall (1980), Small (1988) y Stoyan et al. (1995).

Por otra parte, Christaller (1933) elaboró la *Teoría del Lugar Central*, desarrollada para explicar la distribución espacial de los asentamientos humanos. Según esta teoría, un lugar central sería aquel que pudiese ofrecer servicios de determinadas clase y su área de alcance tendría forma hexagonal, para garantizar la prestación de ese servicio a todo el espacio.

Una forma de comprobar si se cumple la *Teoría del Lugar Central* en una región, es examinar la forma de los triángulos formados por un asentamiento y dos de sus vecinos, y comprobar si son triángulos más o menos equiláteros que lo que se esperaría bajo hipótesis de aleatoriedad. Una triangulación de las ciudades es la triangulación Delaunay, estudiada por Mardia et al. (1977), Green y Sibson (1977) y Mardia (1989). Estos triángulos forman configuraciones de puntos no etiquetadas.

Un ámbito de aplicación donde el problema del emparejamientos de configuraciones no etiquetadas tiene una gran importancia, es el análisis de imágenes donde se han desarrollado varios métodos algorítmicos de emparejamientos.

Rangarajan et al. (1997), extendieron la técnica Procrustes a configuraciones no

etiquetadas (*softassign procrustes algorithm*) y la aplicaron a imágenes digitalizadas del cerebro de primates.

También el emparejamiento de dos imágenes es de gran interés en las investigaciones en *visión computarizada*, también conocida como *reconocimiento de objetos*, que se basa en la construcción de sistemas artificiales que obtienen información a partir de imágenes, como secuencias de vídeos o fotografías desde distintas cámaras.

Gold et al. (1998) desarrollaron un algoritmo rápido y robusto (*DAS, Deterministic Annealing and Softassignment*) que consiste en encontrar una función afín entre las configuraciones de puntos que representan las imágenes y una matriz de correspondencias, a través de un problema de minimización con restricciones.

Para ello se apoyaron en la técnica denominada *softassign* (basada en el trabajo de Kosowsky y Yuille, 1994) y el método de simulación de *annealing*. Las restricciones recogen el hecho de que las filas y las columnas de dicha matriz de correspondencias deben sumar uno y además, estar formada por unos o ceros (en el caso de ser cuadrada, resulta ser una matriz permutación). En general, la simulación annealing encuentra una buena aproximación a un mínimo global de una función dada en un espacio grande de búsqueda y suele usarse cuando éste es discreto. La matriz de correspondencias se encuentra como límite de matrices estocásticas, es decir, matrices cuyos elementos son no negativos y sus filas y columnas suman uno. De forma intuitiva, puede interpretarse como el caso continuo de una matriz permutación.

El problema de la modelización de los emparejamientos de configuraciones de distinto número de puntos, se resuelve definiendo en la matriz de correspondencias una nueva fila y columna que identifique los no emparejamientos, o outliers.

Liu et al. (2004) realizaron una extensión del algoritmo *DAS* incluyendo modificaciones en la función objetivo a minimizar.

Otros algoritmos de búsqueda de emparejamientos son los métodos *tree-pruning*, con los que se representan inicialmente todos los posibles emparejamientos en forma

de árbol, y se van *talando* sus ramas, según van eliminándose los emparejamientos menos probables (Baird (1984), Grimson y Lozano-Pérez (1987) y Umeyama (1993)).

Cross y Hancock (1998), resolvieron el problema utilizando el algoritmo *EM* en conjunto de puntos en dos dimensiones. Walker (1999) adaptó su método al caso de emparejamiento en electroforesis en gel, técnicas empleadas para separar moléculas basándose en propiedades como el tamaño o la forma.

En el ámbito de la Biología, Taylor et al. (2003) y Kent et al. (2004) estudiaron el emparejamiento de moléculas de proteínas utilizando también el algoritmo *EM* para la estimación, definiendo los emparejamientos entre dos configuraciones a través del conjunto de permutaciones entre los dos conjuntos de puntos.

A continuación, se presentan dos publicaciones recientes sobre el problema de emparejamiento de configuraciones no etiquetadas utilizando metodología bayesiana. En ambas, se definen los emparejamientos entre pares de configuraciones a través de una matriz de correspondencias, que se convierte en un parámetro más de los modelos.

Dryden et al. (2007) comparan configuraciones no etiquetadas y consideran una aplicación con moléculas cuyos átomos son identificados con los puntos de estas configuraciones. También incluyen en el modelo información adicional sobre los átomos, en particular, la carga del mismo y el radio van der Waals que ayudan a estimar de forma óptima los emparejamientos.

La importancia de la comparación de moléculas activas, con características conocidas, con moléculas desconocidas, es que puede ayudar a encontrar similitudes entre ambas. Si se consigue encontrar un emparejamiento estructural entre ambos tipos de moléculas, es posible deducir la funcionalidad de las moléculas desconocidas.

Por lo general, las coordenadas de las moléculas se representan sin tener en cuenta su localización y su orientación, de manera que se comparan prescindiendo de dónde estén situadas y si están o no giradas. Es por ello por lo que el modelo

planteado asume que la verosimilitud es invariante ante rotaciones y traslaciones de los datos, de modo que el problema se centra en encontrar los emparejamientos entre los átomos de las moléculas, es decir, una matriz de etiquetas que identifique los emparejamientos.

En el trabajo de Green y Mardia (2006) se aborda el problema del emparejamiento de dos configuraciones de puntos no etiquetados, o parcialmente etiquetados, identificando algunos emparejamientos y sus aplicaciones en conformación de moléculas de proteínas. Se plantea un modelo bayesiano jerárquico donde se supone la existencia de una configuración de puntos de referencia desconocida proveniente de un Proceso de Poisson. Los autores realizan inferencias bayesianas sobre los emparejamientos, al mismo tiempo que estiman las transformaciones afines que definen las superposiciones, entendidas como transformaciones entre los sistemas de coordenadas donde se representan las dos configuraciones.

Existen dos diferencias con respecto al enfoque hecho por Dryden et al. (2007). La primera es que el modelo supone que las dos configuraciones se superponen a una configuración de referencia, es decir, son *generadas* a partir de los puntos del Proceso de Poisson de esta configuración de referencia y, así, proponen un modelo simétrico para ambas configuraciones. La segunda diferencia es que los parámetros de las transformaciones que definen la superposición son incluidos en el modelo.

Evans y Dryden (2008) abordan la comparación de los modelos presentados en los trabajos de Dryden et al. (2007) y de Green y Mardia (2006) y también mejoran la convergencia del algoritmo utilizado en Dryden et al. (2007), proponiendo saltos más grandes en la fase de calentamiento del algoritmo MCMC, usado en la estimación de los parámetros.

Capítulo 2

Definiciones y conceptos previos

2.1. Introducción

El objetivo de este segundo capítulo es presentar varios temas fundamentales en los que se basan los modelos estudiados en esta tesis.

En primer lugar, se exponen los aspectos más importantes del proceso de Poisson espacial homogéneo, en particular en \mathbb{R}^2 , ya que uno de los modelos de emparejamiento entre configuraciones que se estudiará en los capítulos 3 y 4, está basado en este tipo de proceso puntual.

A continuación se presentan las principales transformaciones matriciales de configuraciones de puntos en \mathbb{R}^d , estudiando desde las transformaciones más simples hasta composiciones de ellas que dan lugar a transformaciones lineales más generales, todo ello ilustrado a través de representaciones geométricas.

En tercer lugar se realiza un análisis del problema del emparejamiento de dos configuraciones no etiquetadas asumiendo una transformación afín entre ellas, modelo que será el origen de la generalización a más de dos configuraciones y que será la base del capítulo 3.

Como introducción a las transformaciones no lineales entre configuraciones que se estudiarán en el capítulo 4, se realiza una introducción a las redes neuronales desde la perspectiva de los modelos de regresión no lineales.

Por último, y como presentación de la técnica multivariante utilizada en las aplicaciones en Bioinformática realizadas en este trabajo, se resumen los conceptos claves de la técnica multivariante INDSCAL (INDividual differences SCALing), modelo de escalamiento multidimensional con ponderaciones.

2.2. Proceso de Poisson espacial

2.2.1. Introducción

La disposición de puntos en una región del espacio, sea éste \mathbb{R} , \mathbb{R}^2 ó en general \mathbb{R}^n , puede presentar distintos patrones. Si la nube de puntos presenta un aspecto homogéneo sobre la región, se dice que sigue un modelo regular. Por el contrario, si los puntos presentan ciertos agrupamientos se dice que siguen un modelo agregado. En las figuras 2.1 y 2.2 se muestran dos ejemplos de este tipo de patrones.

Mediante el proceso de Poisson espacial se modelizan disposiciones de puntos en el espacio que son aleatorias, y por tanto se dice que siguen un modelo aleatorio. Este proceso es la base de la teoría de procesos puntuales espaciales y modeliza el mecanismo aleatorio más sencillo para generar patrones aleatorios de puntos. En la figura 1.3 se muestra una nube de puntos de un modelo aleatorio.

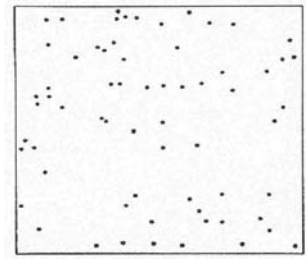
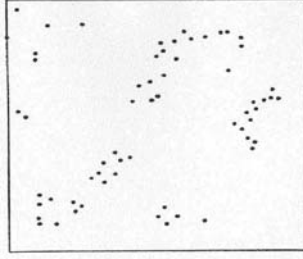
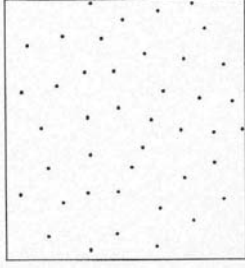


Figura 2.1 Modelo regular Figura 2.2 Modelo agregado Figura 2.3 Modelo aleatorio

Sin pérdida de generalidad, nos centraremos en el caso de \mathbb{R}^2 . Denotaremos por $N(A)$ el número de eventos o puntos en una región A del plano, $|A|$ al área de A y $d\mathbf{x}$ una región infinitesimal que contiene a \mathbf{x} .

2.2.2. Los procesos puntuales espaciales

Definición 6 Se llama *proceso puntual espacial en \mathbb{R}^2* , a cualquier mecanismo estocástico que genera un conjunto de eventos o puntos \mathbf{x}_i , $i = 1, \dots, n$, situados en el plano.

Definición 7 Se dice que un proceso puntual espacial es *estacionario*, si todas sus propiedades probabilísticas en cualquier región A del plano, se mantienen invariantes bajo traslaciones arbitrarias de la región A . Si además esta invarianza se mantiene bajo rotaciones de A , se dice que el proceso es *isotrópico*.

Definición 8 Sea N un proceso puntual en el plano. Se define la *función de intensidad de primer orden del proceso* como,

$$\lambda(\mathbf{x}) = \lim_{|d\mathbf{x}| \rightarrow 0} \left\{ \frac{E[N(d\mathbf{x})]}{|d\mathbf{x}|} \right\}.$$

En un proceso estacionario $\lambda(\mathbf{x})$ es constante e igual a λ para todo \mathbf{x} , y representa el número medio de eventos por unidad de área.

Definición 9 *Se define la función de intensidad de segundo orden del proceso N como*

$$\lambda_2(\mathbf{x}, \mathbf{y}) = \lim_{\substack{|\mathbf{dx}| \rightarrow 0 \\ |\mathbf{dy}| \rightarrow 0}} \left\{ \frac{E[N(\mathbf{dx})N(\mathbf{dy})]}{|\mathbf{dx}| |\mathbf{dy}|} \right\}.$$

En un proceso estacionario, $\lambda_2(\mathbf{x}, \mathbf{y}) = \lambda_2(\mathbf{x} - \mathbf{y})$ es decir, es función de la diferencia $\mathbf{x} - \mathbf{y}$.

Definición 10 *Para n puntos en una región A , se define la distancia al vecino más próximo desde el punto i como la distancia desde el i -ésimo punto al punto más cercano en A .*

La función de distribución empírica de la variable Y , distancia de un punto al vecino más próximo, es

$$\hat{G}(y) = \frac{\#(y_i \leq y)}{n},$$

donde con $\#$ se denota el número de puntos.

Esta variable tiene interés en situaciones donde los eventos o puntos *compiten* entre ellos y por lo tanto es necesario fijar una distancia mínima. Tal puede ser el caso de los árboles en una región, donde la necesidad de nutrientes en la tierra y de la luz solar puede convertirlos en competidores.

Este concepto puede generalizarse definiendo la distancia al k -ésimo vecino más próximo desde el punto i , como la distancia desde el i -ésimo punto al k -ésimo más cercano que se encuentra en A .

2.2.3. El Proceso de Poisson espacial

Definición 11 *El proceso de conteo N es un proceso de Poisson homogéneo de intensidad $\lambda > 0$, si verifica las siguientes propiedades:*

- (i) Para cualquier región finita del plano A , $N(A)$ sigue una distribución de Poisson de media $\lambda |A|$.
- (ii) Para toda colección $\{A_1, \dots, A_m\}$ de conjuntos disjuntos, las variables aleatorias $N(A_1), \dots, N(A_m)$ son independientes.

Proposición 1 Sea A una región del plano con $|A| > 0$ con exactamente un punto. Entonces, la probabilidad de que el punto se encuentre en $B \subset A$ es

$$P\{N(B) = 1 \mid N(A) = 1\} = \frac{|B|}{|A|}.$$

Demostración. $A = B \cup B^c$. Por (ii) $N(B)$ y $N(B^c)$ son independientes y por (i) $N(B)$ y $N(B^c)$ siguen una distribución Poisson de medias $\lambda |B|$ y $\lambda |B^c|$, respectivamente. Así,

$$\begin{aligned} P\{N(B) = 1 \mid N(A) = 1\} &= \frac{P\{N(B) = 1, N(B^c) = 0\}}{P\{N(A) = 1\}} \\ &= \frac{P\{N(B) = 1\}P\{N(B^c) = 0\}}{P\{N(A) = 1\}} \\ &= \frac{[(\lambda |B|)^1 e^{-\lambda |B|}/1!][(\lambda |B^c|)^0 e^{-\lambda |B^c|}/0!]}{(\lambda |A|)^1 e^{-\lambda |A|}/1!} = \frac{|B|}{|A|}. \end{aligned}$$

■

La generalización de este resultado a más de un punto, se recoge en la siguiente proposición.

Proposición 2 Sea A una región del plano con $|A| > 0$ con exactamente n puntos. Entonces para toda partición disjunta de A , $A_1 \cup \dots \cup A_m = A$, y para todo $k_1, \dots, k_m \in \mathbb{R}$ tales que $\sum_{i=1}^m k_i = n$ se tiene que

$$P(N(A_1) = k_1, \dots, N(A_m) = k_m \mid N(A) = n) = \frac{n!}{k_1! \dots k_m!} \left(\frac{|A_1|}{|A|} \right)^{k_1} \dots \left(\frac{|A_m|}{|A|} \right)^{k_m}.$$

Demostración. Para todo $i = 1, \dots, m$, por la definición 11(ii) $N(A_i)$ son independientes y por definición 11(i) $N(A_i)$ siguen una distribución Poisson de media $\lambda |A_i|$. Además, por la proposición 1, la probabilidad de que un punto pertenezca al conjunto A_i es $\frac{|A_i|}{|A|}$. Por lo tanto, para todo $k_1, \dots, k_m \in \mathbb{R}$ tales que $\sum_{i=1}^m k_i = n$, a partir de la distribución multinomial, se tiene que

$$P(N(A_1) = k_1, \dots, N(A_m) = k_m) \mid N(A) = n = \frac{n!}{k_1! \dots k_m!} \left(\frac{|A_1|}{|A|} \right)^{k_1} \dots \left(\frac{|A_m|}{|A|} \right)^{k_m}.$$

■

Este resultado asegura que los n eventos en A se distribuyen uniformemente sobre A .

Observación 1 *La definición del proceso de Poisson espacial implica unas propiedades que verifica el proceso puntual, análogas al caso unidimensional y que son las siguientes:*

1. *La distribución de $N(A)$ depende de la región A sólo a través de su tamaño $|A|$ con la propiedad que*

$$P\{N(A) \geq 1\} = \lambda |A| + o(|A|) \text{ cuando } |A| \downarrow 0.$$

2. *Para $m = 2, 3, \dots$, si A_1, \dots, A_m son regiones disjuntas, entonces $N(A_1), \dots, N(A_m)$ son variables aleatorias independientes y $N(A_1 \cup \dots \cup A_m) = N(A_1) + \dots + N(A_m)$.*

- 3.

$$\lim_{|A| \rightarrow 0} \frac{P\{N(A) \geq 1\}}{P\{N(A) = 1\}} = 1.$$

La propiedad 1 indica que la distribución de $N(A)$ no depende de la forma ni de la localización de A , sino sólo de su tamaño $|A|$. La propiedad 2 implica que un punto en una región ni influye ni es influido por la presencia de otro punto en una segunda región que no se solape con la primera. Y, por último, la tercera propiedad

establece el hecho de que los eventos se presentan de uno en uno, no siendo posible el solapamiento de dos en una misma localización.

La función de intensidad del proceso de Poisson homogéneo es constante e igual a λ , ya que

$$\lambda(\mathbf{x}) = \lim_{|d\mathbf{x}| \rightarrow 0} \left\{ \frac{E[N(d\mathbf{x})]}{|d\mathbf{x}|} \right\} = \lim_{|d\mathbf{x}| \rightarrow 0} \left\{ \frac{\lambda |d\mathbf{x}|}{|d\mathbf{x}|} \right\} = \lambda,$$

y la función de intensidad de segundo orden es λ^2 , ya que se tiene que

$$\begin{aligned} \lambda_2(\mathbf{x}, \mathbf{y}) &= \lim_{\substack{|d\mathbf{x}| \rightarrow 0 \\ |d\mathbf{y}| \rightarrow 0}} \left\{ \frac{E[N(d\mathbf{x})N(d\mathbf{y})]}{|d\mathbf{x}| |d\mathbf{y}|} \right\} = \lim_{\substack{|d\mathbf{x}| \rightarrow 0 \\ |d\mathbf{y}| \rightarrow 0}} \left\{ \frac{E[N(d\mathbf{x})] \cdot E[N(d\mathbf{y})]}{|d\mathbf{x}| |d\mathbf{y}|} \right\} \\ &= \lim_{\substack{|d\mathbf{x}| \rightarrow 0 \\ |d\mathbf{y}| \rightarrow 0}} \left\{ \frac{\lambda |d\mathbf{x}| \cdot \lambda |d\mathbf{y}|}{|d\mathbf{x}| |d\mathbf{y}|} \right\} = \lambda^2. \end{aligned}$$

Con respecto a la variable Y , distancia de un punto al vecino más próximo, denotamos por Y_i a la distancia desde el punto i al vecino más próximo. Así la probabilidad de que se encuentre a menos de una distancia y será:

$$P(Y_i \leq y) = \frac{\text{Área del círculo de radio } y}{|A|} = \frac{\pi y^2}{|A|}.$$

Por lo tanto la función de distribución de Y es,

$$\begin{aligned} G(y) &= P\{Y \leq y\} \\ &= 1 - P\{\text{todos los puntos están a una distancia mayor que } y\} \\ &= 1 - \prod_{j=1}^{n-1} P\{Y_j > y\} = 1 - \left(1 - \frac{\pi y^2}{|A|}\right)^{n-1}. \end{aligned}$$

Una aproximación para n grande, tomando $\lambda = n/|A|$ es entonces,

$$G(y) = 1 - \exp(-\lambda \pi y^2) \quad y \geq 0.$$

Por tanto, la función de densidad de Y es

$$f(y) = 2\lambda \pi y \exp(-\lambda \pi y^2).$$

Un tipo de proceso puntual no estacionario es el proceso de Poisson no homogéneo donde la función de intensidad no es constante, sino es función de \mathbf{x} .

Definición 12 *El proceso de conteo N es un proceso de Poisson no homogéneo con función de intensidad $\lambda(\mathbf{x})$, si verifica las siguientes propiedades:*

- (i) *Para cualquier región finita del plano A , $N(A)$ sigue una distribución de Poisson de media $\int_A \lambda(\mathbf{x})d\mathbf{x}$.*
- (ii) *Dado $N(A) = n$, los n eventos en A forman una muestra aleatoria simple de la distribución sobre A , cuya función de distribución es proporcional a $\lambda(\mathbf{x})$.*

Simulación de un proceso de Poisson espacial

Si se condiciona a un valor fijo de $N(A)$, un método directo para generar un proceso de Poisson homogéneo en A sería distribuir los sucesos independientemente según la distribución uniforme sobre A . Si la forma de A no es sencilla, se puede simular sobre una región de una forma más sencilla (rectángulo o círculo, por ejemplo) donde esté incluida la región A , y prescindir de aquellos eventos que no han caído en A . Hsuan (1979), proporciona un algoritmo para generar puntos uniformemente distribuidos sobre un polígono cualquiera.

Por otro lado, si se requiere que $N(A)$ sea aleatorio, se puede aplicar este mismo método previa simulación de $N(A)$ a partir de la distribución de Poisson adecuada.

Ross (1997) presenta la simulación de un proceso de Poisson homogéneo en un círculo de radio fijo r y centro el origen, a partir de las coordenadas polares de los puntos. Denotamos por $C(r)$ al círculo centrado en el origen y de radio r . Por definición 11(i), el número de puntos en $C(r)$ tiene una distribución Poisson de parámetro $\lambda\pi r^2$. Sea R_i la distancia del origen a su i -ésimo vecino más próximo.

Entonces,

$$\begin{aligned} P\{\pi R_1^2 > x\} &= P\{R_1 > \sqrt{\frac{x}{\pi}}\} = P\{\text{Ningún punto se encuentre en } C(\sqrt{\frac{x}{\pi}})\} \\ &= P\{N(C(\sqrt{\frac{x}{\pi}})) = 0\} = \exp\{-\lambda x\}. \end{aligned}$$

Por lo tanto, el área del círculo centrado en el origen y de radio R_1 , (distancia del origen al vecino más próximo), sigue una distribución exponencial de parámetro λ .

Este razonamiento puede repetirse para obtener el siguiente resultado general, que afirma que las áreas de las coronas circulares formadas como diferencia de dos círculos de radio R_i y R_{i-1} $i = 1, \dots, n$, son variables independientes y siguen una distribución exponencial de parámetro λ .

Proposición 3 *Tomando $R_0 = 0$, se tiene que para todo $i \geq 1$, $\pi R_i^2 - \pi R_{i-1}^2$ son variables aleatorias independientes de media λ .*

Además, por simetría, se tiene que los respectivos ángulos de los puntos del proceso de Poisson son independientes y se distribuyen uniformemente sobre $(0, 2\pi)$, por lo que el algoritmo que propone Ross (1997) para simular un proceso de Poisson sobre un círculo de radio r y centro el origen, es el siguiente:

Algoritmo 13

Paso 1 *Simular exponenciales independientes de tasa λ , X_1, X_2, \dots, X_{N-1} donde*

$$N = \text{Min}\{n : X_1 + \dots + X_n > \pi r^2\}.$$

Paso 2 *Si $N = 1$, parar, y entonces no hay puntos en $C(r)$.*

En otro caso, para $i = 1, \dots, N - 1$, tomar

$$R_i = \sqrt{\frac{X_1 + \dots + X_i}{\pi}},$$

es decir, $\pi R_i^2 = X_1 + \dots + X_i$.

Paso 3 Simular U_1, \dots, U_{N-1} de una distribución $U(0, 1)$.

Paso 4 Las coordenadas polares de los $N - 1$ puntos del proceso de Poisson de tasa λ son

$$(R_i, 2\pi U_i), \quad i = 1, \dots, N - 1.$$

Ross (1997) también aplica el mismo razonamiento a regiones del plano generales, limitadas por el eje x (intervalo $(0, T)$) y una función positiva $f(x)$. El caso particular de $f(x) = k$, equivale a simular en el rectángulo $(0, T) \times (0, k)$.

Por último, para simular un proceso de Poisson no homogéneo, Lewis y Shedler (1979) sugieren un algoritmo basado en el método de rechazo. Básicamente consiste en simular un proceso de Poisson homogéneo sobre A con intensidad $\lambda_0 = \max_{\mathbf{x} \in A} \lambda(\mathbf{x})$ y elegir el evento en \mathbf{x} con probabilidad $\lambda(\mathbf{x})/\lambda_0$.

2.3. Transformaciones matriciales

Uno de los aspectos más relevantes de este trabajo se refiere a la representación de configuraciones de puntos, su emparejamiento y su relación mediante distinto tipo de transformaciones, en particular, transformaciones lineales. En esta sección se presentan distintos tipos de transformaciones matriciales en \mathbb{R}^d .

2.3.1. Transformaciones matriciales simples

Definición 14 Se llama configuración de puntos a un conjunto de k puntos en \mathbb{R}^d , de manera que se representa a través de la matriz X , de dimensiones $k \times d$, que recoge las coordenadas cartesianas de los k puntos en \mathbb{R}^d .

Definición 15 Una transformación $T : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ es lineal si verifica las siguientes propiedades:

1. Para todo $\mathbf{x} \in \mathbb{R}^d$ existe un único vector $\mathbf{x}^* = T(\mathbf{x}) \in \mathbb{R}^d$.
2. Para todo $k \in \mathbb{R}$ y todo $\mathbf{x} \in \mathbb{R}^d$, se tiene que $T(k\mathbf{x}) = kT(\mathbf{x})$.
3. Para todo $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ se tiene que $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$.

Cualquier transformación lineal de \mathbb{R}^d en \mathbb{R}^d , puede representarse mediante una matriz cuadrada no singular A de dimensiones $d \times d$, es decir, mediante una transformación matricial general,

$$T(\mathbf{x}) = A\mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^d.$$

Observación 2 En general, si se quiere aplicar una transformación matricial a una configuración de puntos X , al recogerse los puntos en vectores filas, la configuración transformada se expresa multiplicando por la derecha por la matriz traspuesta, es decir,

$$X^* = T(X) = XA^T.$$

Definición 16 Una matriz cuadrada $A_{n \times n}$ es ortogonal si verifica

$$A^T A = I_n,$$

es decir, cualquier par de vectores fila o vectores columna son ortogonales y cada vector tiene módulo 1.

Una matriz ortogonal verifica las siguiente propiedades:

1. $A^{-1} = A^T$.
2. $|A| = 1$ ó $|A| = -1$.
3. $C = AB$ es ortogonal si A y B son ortogonales.

Las matrices ortogonales pueden utilizarse para representar cambios de base o rotaciones.

A continuación se presentan las siguientes transformaciones simples: rotación, traslación, reflexión, dilatación y estiramiento.

Rotación

Definición 17 Una matriz $\Gamma_{n \times n}$ es una matriz de rotación si es ortogonal ($\Gamma^T \Gamma = \Gamma \Gamma^T = I_n$) y $|\Gamma| = +1$. También se le conoce como rotación propia.

Las rotaciones son un tipo particular de transformación lineal y juegan un papel muy importante en el análisis de transformaciones matriciales generales.

Se puede hablar de rotaciones desde dos puntos de vista:

1. *Rotaciones de puntos*, donde el sistema de referencia se mantiene fijo y lo que se rotan son los puntos con respecto al origen, según el sentido de las manecillas del reloj (ángulos negativos) o en sentido contrario (ángulos positivos).
2. *Rotaciones de ejes*, donde los puntos se mantienen fijos y lo que se rota es el sistema de referencia (origen y base) según también el sentido de las manecillas del reloj o en sentido contrario. En este caso, los puntos son expresados en términos del nuevo sistema de referencia rotado.

Aunque ambos tipos de rotación son equivalentes, la rotación de puntos es más sencilla de utilizar aunque la segunda interpretación es muy empleada en el desarrollo de las técnicas de análisis multivariante. Las coordenadas de un punto al que se le aplica una rotación de ángulo $-\alpha$ coinciden con las coordenadas del mismo punto con respecto al eje rotado un ángulo $+\alpha$. Este resultado se ilustra en las figuras 2.4 y 2.5, donde se representa el punto $\mathbf{x} = (1, 2)$ respecto al eje de coordenadas rotado

30° en sentido positivo (figura 2.4), y las coordenadas del mismo punto, rotado 30° negativo (figura 2.5).

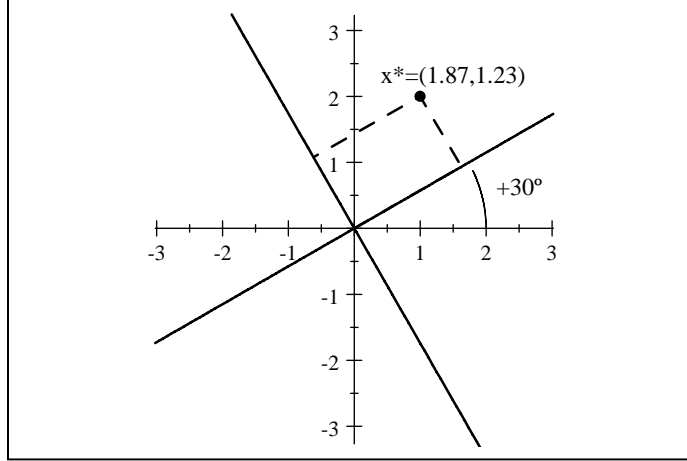


Figura 2.4 Rotación de ejes en sentido positivo

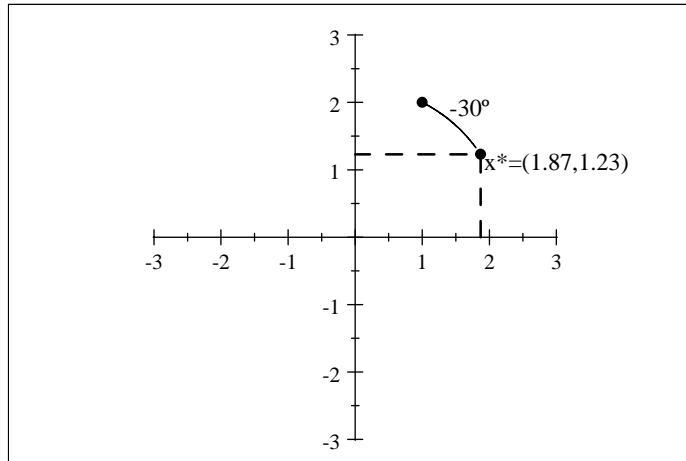


Figura 2.5 Rotación de punto en sentido negativo

Observación 3 *En el caso de dos dimensiones, la matriz de rotación con la que se rota a un punto \mathbf{x} con un ángulo α en sentido positivo, es de la forma*

$$\Gamma = \begin{pmatrix} \cos \alpha & -\operatorname{sen} \alpha \\ \operatorname{sen} \alpha & \cos \alpha \end{pmatrix},$$

y en el caso de que la rotación sea en sentido de las agujas del reloj, utilizando las propiedades trigonométricas $\operatorname{sen} \alpha = -\operatorname{sen}(-\alpha)$ y $\cos \alpha = \cos(-\alpha)$, la matriz de

rotación será

$$\Gamma^* = \begin{pmatrix} \cos \alpha & \operatorname{sen} \alpha \\ -\operatorname{sen} \alpha & \cos \alpha \end{pmatrix} = \Gamma^T.$$

Así, el punto $\mathbf{x}^T = (x_1, x_2)$ rotado en sentido positivo α , tendrá por coordenadas

$$\mathbf{x}^* = \begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} = \Gamma \mathbf{x} = \begin{pmatrix} x_1 \cos \alpha - x_2 \operatorname{sen} \alpha \\ x_1 \operatorname{sen} \alpha + x_2 \cos \alpha \end{pmatrix}.$$

y en sentido negativo α ,

$$\mathbf{x}^* = \begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} = \Gamma^* \mathbf{x} = \begin{pmatrix} x_1 \cos \alpha + x_2 \operatorname{sen} \alpha \\ -x_1 \operatorname{sen} \alpha + x_2 \cos \alpha \end{pmatrix}.$$

La matriz de rotación utilizada en la figura 1.5 para rotar el punto $\mathbf{x}^T = (1, 2)$ un ángulo de $\alpha = 30^\circ$ en sentido negativo es:

$$\Gamma^* = \begin{pmatrix} 0,87 & 0,5 \\ -0,5 & 0,87 \end{pmatrix}$$

por lo que el punto rotado es

$$\mathbf{x}^* = \Gamma^* \mathbf{x} = \begin{pmatrix} 0,87 + 2 \cdot 0,5 \\ 1(-0,5) + 2 \cdot 0,87 \end{pmatrix} = \begin{pmatrix} 1,87 \\ 1,23 \end{pmatrix}.$$

Observación 4 En general, para rotar en sentido positivo una configuración de k puntos $X_{k \times d}$, la configuración transformada se expresa de la forma

$$X^* = X\Gamma^*$$

donde

$$\Gamma^* = \begin{pmatrix} \cos \alpha & -\operatorname{sen} \alpha \\ \operatorname{sen} \alpha & \cos \alpha \end{pmatrix}^T = \begin{pmatrix} \cos \alpha & \operatorname{sen} \alpha \\ -\operatorname{sen} \alpha & \cos \alpha \end{pmatrix}.$$

Así, la rotación en sentido positivo de una configuración se consigue multiplicando por la derecha la matriz de configuración por la matriz de rotación Γ^* .

Traslación

Definición 18 Una traslación se consigue sumando un vector constante k -dimensional al punto, es decir

$$\mathbf{x}^* = \mathbf{x} + \boldsymbol{\gamma}$$

o bien, a una configuración de puntos,

$$X^* = X + \mathbf{1}_k \gamma^T.$$

Definición 19 Las transformaciones rígidas de una matriz de configuración $X_{k \times m}$ son el conjunto de matrices X rotadas y trasladadas, es decir

$$X^* = X\Gamma + \mathbf{1}_k \gamma^T$$

donde Γ es una matriz de rotación, $\mathbf{1}_k$ es un vector columna de unos de dimensión k y γ es el vector de traslación.

Estas transformaciones se caracterizan porque preservan los ángulos entre vectores, longitudes y distancias entre puntos.

Reflexión

Esta transformación consiste en multiplicar un número impar de coordenadas de los puntos por -1 . Así, por ejemplo en \mathbb{R}^2 , la matriz de reflexión que cambia de signo a las primeras coordenadas de los puntos sería

$$A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

En las figuras 2.6 y 2.7 se representan una configuración de 9 puntos, y la configuración reflejada a través de esta matriz, respectivamente.

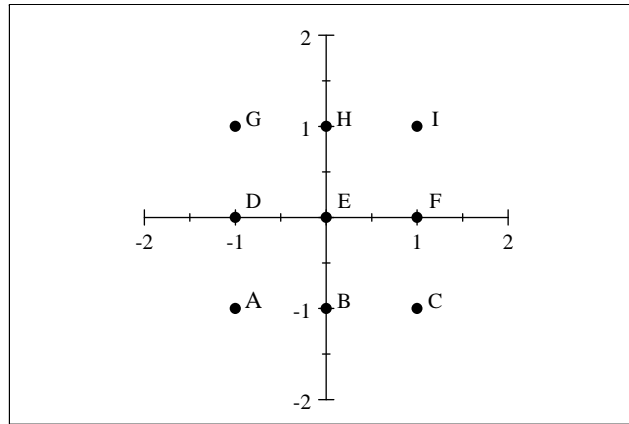


Figura 2.6 Configuración original

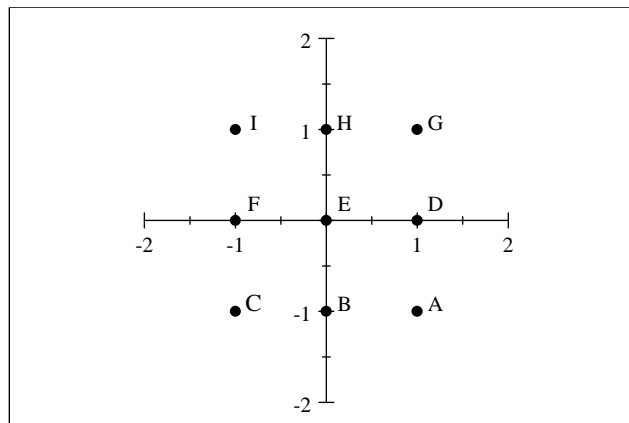


Figura 2.7 Configuración reflejada

Observación 5 Las matrices ortogonales con $|\Gamma| = -1$, se les llama *impropias* y representan una rotación seguida de un número impar de reflexiones de los ejes, es decir, en el caso de \mathbb{R}^2 , rotación más una reflexión, en el caso de \mathbb{R}^3 , rotación más una o tres reflexiones, y así sucesivamente.

Estas matrices se obtienen a partir de las matrices ortogonales propias ($|\Gamma| = +1$), multiplicando por -1 un número impar de filas. Esto hace que $|\Gamma| = -1$.

Dilatación

Una dilatación de un punto \mathbf{x} consiste en multiplicarlo por un escalar s ($\mathbf{x}^* = s\mathbf{x}$), de manera que si $s > 1$ se consigue una extensión uniforme del vector de puntos mientras que si $s < 1$, se consigue una contracción del mismo

Por otro lado la dilatación de una configuración de puntos X , se consigue multiplicando la configuración por una matriz diagonal, con todos los elementos de la misma iguales a s ($X^* = XU = sXI_k = sX$).

Estiramiento

Una transformación de estiramiento, es una generalización de la dilatación, donde las coordenadas de los puntos se dilatan o contraen multiplicándolas por distintos escalares. Así, la matriz involucrada en esta transformación es una matriz diagonal D , con diferentes valores en la misma. Así, por ejemplo, tomando $\mathbf{x} = (-1, -1)$,

$$\mathbf{x}^* = D\mathbf{x} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}.$$

En la figura 2.8 se recoge la configuración X de la figura 2.6, transformada mediante este estiramiento.

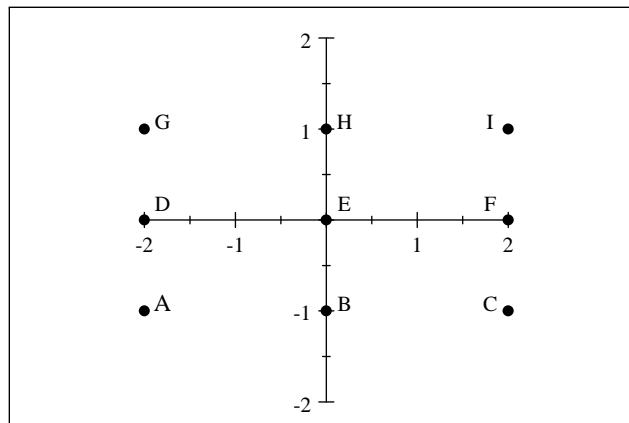


Figura 2.8 Efecto geométrico de una transformación de estiramiento

2.3.2. Composición de transformaciones matriciales simples

Una vez presentadas las transformaciones simples y su interpretación geométrica, en esta sección se muestran transformaciones lineales más generales y complejas, obtenidas como composición de las primeras. En concreto, nos centramos en las transformaciones lineales arbitrarias de matriz no singular A , transformaciones afines y transformaciones de similitud.

Una transformación lineal arbitraria se puede interpretar como una composición de transformaciones más simples, gracias a la descomposición singular de una matriz.

Teorema 20 *Toda matriz $A_{(n \times k)}$ de rango r puede expresarse como*

$$A = UDV^T$$

donde $U_{n \times r}$ y $V_{k \times r}$ son matrices con vectores columnas ortogonales de norma uno y $D_{r \times r}$ es una matriz diagonal con elementos positivos ordenados. La matriz diagonal D contiene las raíces cuadradas de los valores propios no nulos de las matrices AA^T o A^TA , que son positivos. La matriz U contiene en columnas los vectores propios asociados a valores propios no nulos de AA^T y V contiene en columnas los vectores propios asociados a valores propios no nulos de A^TA . Los elementos diagonales de D se denominan los valores singulares de la matriz A .

Corolario 21 *Si $A_{p \times p}$ es una matriz cuadrada no singular, entonces se puede escribir de la forma*

$$A = UDV^T$$

donde $U^TU = UU^T = I$ y $V^TV = VV^T = I$ y D es una matriz diagonal.

Este resultado garantiza que cualquier transformación matricial general no singular con valores reales, puede descomponerse de manera única en el producto de las siguientes transformaciones simples:

1. una rotación, un estiramiento y otra rotación, o bien
2. una rotación, una reflexión, un estiramiento y otra rotación.

Ejemplo 22 Consideramos la transformación matricial general definida con la matriz no singular

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

La descomposición singular de A , es

$$A = UDV^T = \begin{pmatrix} -0,41 & -0,91 \\ -0,91 & 0,414 \end{pmatrix} \begin{pmatrix} 5,47 & 0 \\ 0 & 0,37 \end{pmatrix} \begin{pmatrix} -0,58 & -0,82 \\ 0,82 & -0,58 \end{pmatrix}$$

donde U es ortogonal con $|U| = -1$ y V también con $|V| = 1$. Así, la transformación lineal dada por la matriz A se puede descomponer en una rotación de -66° seguida de una reflexión, transformación recogida con la matriz U , un estiramiento, la matriz D , y una rotación de 125° , matriz V^T .

En las figuras 2.9, 2.10 y 2.11 se representa la descomposición de la transformación general A , de la configuración de la figura 2.6.

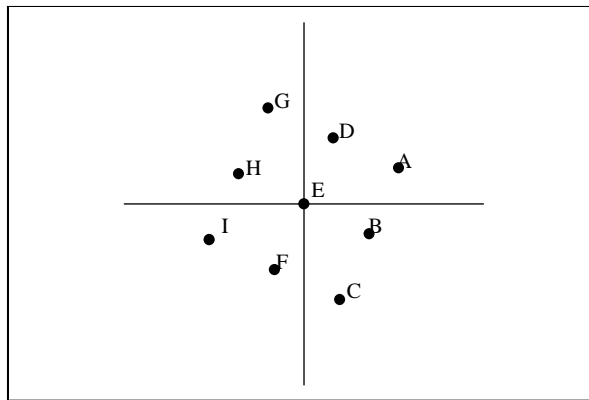


Figura 2.9 Configuración X rotada y reflejada XU

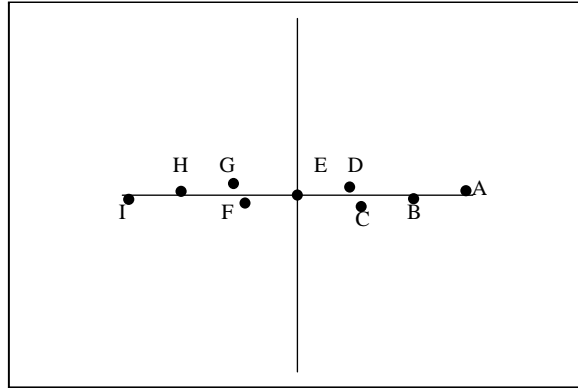


Figura 2.10 Configuración rotada, reflejada y estirada XUD

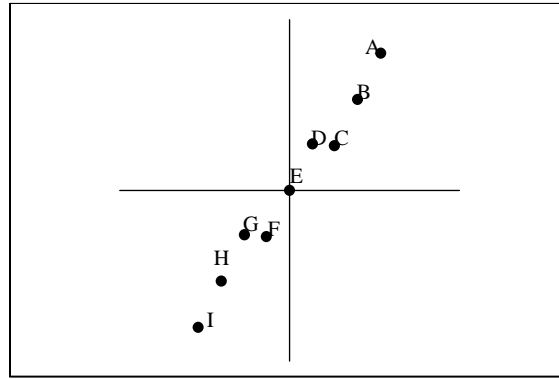


Figura 2.11 Configuración $XA = XUDV^T$

Definición 23 Se define la transformación lineal general afín de un vector \mathbf{x} como

$$\mathbf{x}^* = A\mathbf{x} + \mathbf{c}$$

y de una configuración X como

$$X^* = XA^T + \mathbf{1}\mathbf{c}^T,$$

donde A es una matriz arbitraria cuadrada y \mathbf{c} un vector de \mathbb{R}^d .

Este tipo de transformaciones consisten, pues, en la composición de una transformación lineal general, más una traslación mediante el vector \mathbf{c} . Si no se cambia de origen, entonces estamos ante transformaciones lineales generales homogéneas $\mathbf{x}^* = A\mathbf{x}$.

Definición 24 Una transformación de similitud es aquella que consiste en una rotación, mediante una matriz de rotación propia A , una dilatación de constante s y una traslación de un vector \mathbf{x} , es decir,

$$\mathbf{x}^* = sA\mathbf{x} + \mathbf{c}.$$

En el caso de una configuración X ,

$$X^* = XA^T U + \mathbf{1}\mathbf{c}^T$$

donde U es la matriz escalar $U = sI_2$.

Estas transformaciones se caracterizan porque *estiran* todas las distancias de la configuración por un mismo factor de escala. En el caso de que la transformación deje todas las distancias entre puntos exactamente iguales, se dice que es una *isometría*. Ejemplos de isometría son las traslaciones y las rotaciones.

2.4. El emparejamiento de dos configuraciones de puntos no etiquetadas

En la sección 2.3 se ha representado una configuración de k puntos mediante una matriz X , de dimensiones $k \times d$, que recoge las coordenadas cartesianas de los k puntos en dimensión d . Cuando se procede a estudiar la relación entre dos o más configuraciones es necesario etiquetar los puntos que las definen con el objetivo de identificarlos.

Ya adelantamos en el capítulo 1, que dependiendo del conocimiento que se tenga de las configuraciones que se van a estudiar y de la relación existente entre ellas, la asignación de estas etiquetas puede hacerse de forma aleatoria en cada configuración, de manera que no existe ninguna correspondencia entre los puntos de distintas configuraciones, (configuraciones no etiquetadas), o bien, mediante algún criterio

que permita identificar a puntos de configuraciones distintas con las mismas etiquetas, creando, por lo tanto, emparejamientos entre los puntos de las configuraciones, (configuraciones etiquetadas).

Green y Mardia (2006) abordan el problema del emparejamiento de dos configuraciones de puntos no etiquetados, o parcialmente etiquetados, identificando algunos emparejamientos y asumiendo transformaciones afines entre ellas. Los contenidos del capítulo 3 y parte de los del capítulo 4 de la presente tesis, se han apoyado en este trabajo y consisten en una generalización al caso de más de dos configuraciones y a transformaciones más generales.

2.4.1. El modelo matemático

Sean dos configuraciones de puntos en \mathbb{R}^d , $X = \{x_j, j = 1, \dots, n\}$ e $Y = \{y_k, k = 1, \dots, m\}$ consideradas como observaciones aleatorias de un conjunto de puntos o localizaciones $\{\mu_l\}$ $l = 1, \dots, N$, que definen la configuración de referencia y se desconoce qué puntos de ambas configuraciones se corresponden a cada μ_l . El objetivo es realizar inferencias acerca de los posibles pares (j, k) que se correspondan con la misma localización, lo que indicaría que el punto x_j está emparejado con el punto y_k . Pero además de desconocer los emparejamientos, existe una transformación afín también desconocida entre el espacio de la configuración X y el de la configuración Y , de manera que un y en el espacio de Y , se corresponde con un $x = Ay + \tau$ en el espacio de X . Se supone, sin falta de generalidad, que la configuración de referencia $\{\mu_l \mid l = 1, \dots, N\}$ se localiza en el espacio de X .

Green y Mardia (2006) representan esta situación a través del modelo:

$$\begin{aligned} x_j &= \mu_{\xi_j} + \varepsilon_{1j}, & j &= 1, \dots, n \\ Ay_k + \tau &= \mu_{\eta_k} + \varepsilon_{2k}, & k &= 1, \dots, m \end{aligned} \tag{2.1}$$

donde A y τ , definen la transformación afín entre los dos espacios, ξ_j identifica al punto μ que se corresponde con el punto x_j de la primera configuración, y η_k identifica al punto μ que genera a y_k de la segunda configuración. Además $\{\varepsilon_{1j}\}$ y $\{\varepsilon_{2k}\}$

tienen distribución f_1 y f_2 respectivamente, y son independientes e independientes de $\{\mu_l\}$.

2.4.2. El Proceso de Poisson de la configuración de referencia

Se supone que el conjunto de puntos fijos $\{\mu_l\} \quad l = 1, \dots, N$, forman un proceso de Poisson homogéneo de tasa λ sobre una región $V \subset \mathbb{R}^d$ de volumen v , y que partimos de N observaciones de este proceso en dicha región. Se asume además que cada uno de estos μ_l puede generar de forma independiente, un punto de cada configuración (lo que indicará que son puntos emparejados), sólo un punto de la configuración X , sólo un punto de la configuración Y o ningún punto de ninguna configuración. Así, los $\{\mu_l\}$ quedan clasificados en cuatro clases distintas e independientes, C_{XY} , C_X , C_Y y C_0 , respectivamente. Las probabilidades de que un μ_l pertenezca a cada una de ellas son, $\rho p_X p_Y$, p_X , p_Y y $1 - p_X - p_Y - \rho p_X p_Y$, respectivamente, donde ρ es la tasa de emparejamientos a priori, entendida como una medida a priori de la tendencia a que dos puntos estén emparejados.

Para todo N , n y m , habrá L pares de puntos emparejados en la muestra si y sólo si, hay L puntos μ_l en C_{XY} , $n - L$ puntos en C_X , $m - l$ puntos en C_Y y $N - n - m + L$ puntos en C_0 . Así, los procesos puntuales que cuentan el número de puntos μ_l de cada tipo son procesos de Poisson de tasas $\lambda v \rho p_X p_Y$, $\lambda v p_X$, $\lambda v p_Y$ y $\lambda v (1 - \rho p_X p_Y - p_X - p_Y)$, respectivamente.

2.4.3. La matriz de emparejamientos y su distribución a priori

Con el objetivo de identificar qué puntos de las dos configuraciones están emparejados, Green y Mardia (2006) definen la matriz de emparejamientos M como,

$$M_{jk} = \begin{cases} 1 & \text{si } \xi_j = \eta_k \\ 0 & \text{si } \xi_j \neq \eta_k \end{cases},$$

es decir, $M_{jk} = 1$ indica que x_j e y_k provienen del mismo punto μ_l y por tanto están emparejados y $M_{jk} = 0$ indica que no lo están. Se tiene que $\sum_{j,k} M_{jk} = L$, y que en cada fila y en cada columna de la matriz M hay a lo sumo un 1, es decir, $\sum_j M_{jk} \leq 1$ para todo k , y $\sum_k M_{jk} \leq 1$ para todo j .

Se asume que conocido L , el número de pares de puntos emparejados, M sigue una distribución uniforme, es decir, hay

$$L! \binom{n}{L} \binom{m}{L}$$

posibles matrices de emparejamientos, todas con la misma probabilidad, por lo que

$$p(M|L) = \frac{1}{L! \binom{n}{L} \binom{m}{L}}.$$

Entonces, basándonos en el hecho de que la distribución a priori de L condicionada a n y m es proporcional a

$$\begin{aligned} p(L) &\propto \frac{e^{-\lambda v p_X} (\lambda v p_X)^{n-L}}{(n-L)!} \cdot \frac{e^{-\lambda v p_Y} (\lambda v p_Y)^{m-L}}{(m-L)!} \cdot \frac{e^{-\lambda v \rho p_X p_Y} (\lambda v \rho p_X p_Y)^L}{L!} \\ &\propto \frac{(\rho/\lambda v)^L}{(m-L)!(n-L)!L!}, \end{aligned}$$

donde $L = 0, 1, \dots, \min\{n, m\}$, se llega a que la distribución a priori de M es

$$\begin{aligned} p(M) &= p(L)p(M|L) \propto \frac{(\rho/\lambda v)^L}{(m-L)!(n-L)!L!} \cdot \frac{1}{L! \binom{n}{L} \binom{m}{L}} \\ &\propto (\rho/\lambda v)^L. \end{aligned}$$

2.4.4. Verosimilitud de los datos

Asumiendo la transformación afín entre las configuraciones, Green y Mardia (2006) calculan la expresión de la verosimilitud de los datos que toma la forma,

$$p(x, y|M, A, \tau) \propto v^{-(m+n)} |A|^n \prod_{\{(j,k): M_{jk}=1\}} g(x_j - Ay_k - \tau),$$

donde $g(z) = \int f_1(z+u)f_2(u)du$, es la función de densidad de la diferencia $\varepsilon_{1j} - \varepsilon_{2k}$.

En el caso de asumir que f_1 y f_2 son distribuciones normales, de manera que

$$x_j \sim N_d(\mu_{\xi_j}, \sigma^2 I_d), \quad Ay_k + \tau \sim N_d(\mu_{\eta_k}, \sigma^2 I_d),$$

se tiene que

$$g(z) = \frac{1}{(\sigma\sqrt{2})^d} \varphi\{z/(\sigma\sqrt{2})\},$$

por lo que la verosimilitud de los datos tiene la forma

$$p(x, y | M, A, \tau) \propto |A|^n \prod_{\{(j,k): M_{jk}=1\}} \frac{\rho \varphi_d\{(x_j - Ay_k - \tau)/(\sigma\sqrt{2})\}}{\lambda(\sigma\sqrt{2})^d},$$

donde $\varphi_d(z)$ es la función de densidad de la distribución $N_d(0, I)$.

2.4.5. Inferencias sobre los parámetros

Asumiendo ρ y λ fijos, se puede realizar inferencias sobre el resto de los parámetros, M , τ , σ^2 y A , dados los datos $\{x_j\}$ e $\{y_k\}$.

Se considera el caso en el que A es una matriz de rotación, es decir, es una matriz ortogonal con determinante positivo, ($A^T A = I$ y $|A| = +1$). En este caso, la expresión de la distribución a posteriori de A , condicionada por el resto de los parámetros es

$$P(A | M, \tau, \sigma, x, y) \propto p(A) \exp \left[\text{traza} \left\{ \frac{1}{2\sigma^2} \sum_{\{(j,k): M_{jk}=1\}} y_k (x_j - \tau)^T A \right\} \right].$$

Asumiendo que A tiene una distribución a priori von Mises (véase e.g. Mardia y Jupp, 2000), cuya expresión es

$$p(A) \propto \exp\{\text{traza}(F_0^T A)\},$$

donde F_0 es una matriz fijada, la distribución a posteriori es también una von Mises cuyo parámetro es

$$F = F_0 + \frac{1}{2\sigma^2} \sum_{\{(j,k): M_{jk}=1\}} (x_j - \tau) y_k^T.$$

Green y Mardia (2006) consideran el caso bidimensional, con

$$A = \begin{pmatrix} \cos \theta & -\operatorname{sen} \theta \\ \operatorname{sen} \theta & \cos \theta \end{pmatrix},$$

y también el caso en tres dimensiones donde la matriz de rotación es el producto de matrices de rotaciones elementales,

$$A = A_{12}(\theta_{12})A_{13}(\theta_{13})A_{23}(\theta_{23})$$

donde, para $i < j$, $A_{ij}(\theta_{ij})$ es la matriz de dimensión 3×3 con $m_{ii} = m_{jj} = \cos \theta_{ij}$, $-m_{ij} = m_{ji} = \sin \theta_{ij}$, $m_{rr} = 1$ con $r \neq i, j$ y el resto de los elementos iguales a cero (Khatri y Mardia, 1977).

Por otro lado, las distribuciones a priori que se asumen para τ y para σ^2 son las distribuciones normal y gamma inversa, respectivamente, que son también conjugadas con respecto a errores normales. La estimación de la distribución a posteriori de los tres parámetros A , τ y σ^2 se realiza con saltos Metropolis.

En el caso de emparejamientos desconocidos, la estimación de la matriz de emparejamientos M , se lleva a cabo mediante el método de Metropolis-Hasting, definiendo la distribución propuesta como sigue: en primer lugar, se selecciona al azar uno de los $n + m$ puntos de las dos configuraciones (por ejemplo y sin falta de generalidad x_j). Si x_j está emparejado, con probabilidad p^* se propone eliminar el emparejamiento y con probabilidad $1 - p^*$ se propone cambiar el emparejamiento de y_k a $y_{k'}$. Por el contrario, si x_j no está emparejado, se propone emparejarlo eligiendo de forma aleatoria entre los y_k que no están emparejados. Por último, una vez generada una muestra a posteriori de matrices de emparejamientos, Green y Mardia (2006) proponen como estimador de la matriz de emparejamiento aquella que consigue minimizar cierta función de pérdida.

2.5. Introducción a los modelos estadísticos de redes neuronales artificiales

2.5.1. El modelo biológico

Las redes neuronales artificiales surgieron como modelo de representación y simulación del proceso de aprendizaje del cerebro basado en la estructura de conexión de las neuronas.

El cerebro humano está formado por miles de millones de neuronas que procesan la información. Cada una de ellas trabaja como un simple procesador pero sus múltiples conexiones son las que hacen posible las grandes capacidades de nuestro cerebro.

Una neurona del cerebro consta de un cuerpo celular o *soma*, una o varias prolongaciones cortas que generalmente transmiten impulsos hacia el soma, las *dendritas*, y una prolongación larga, denominada *axón*, que conduce los impulsos desde el soma hacia otra neurona. En la figura 2.12 se presenta un esquema de esta estructura.

La información pasa de neurona a neurona en forma de estímulo eléctrico a través de las dendritas. Cuando se alcanza un nivel de excitación y supera cierto umbral, se envía información a través del axón, en cuyo caso se dice que la neurona está *activada*. Si no se llega a este umbral, la neurona se dice que está *inhibida*.

Esta estructura y este mecanismo de funcionamiento de una neurona del cerebro intenta representarse con las redes neuronales artificiales.

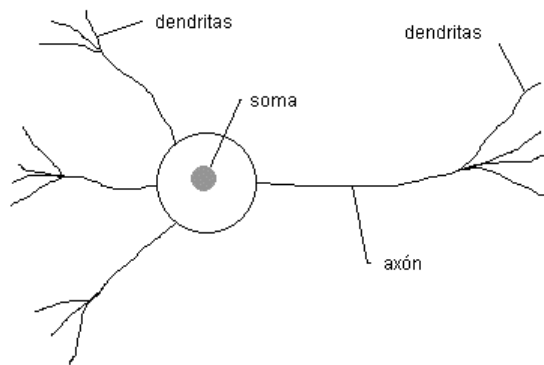


Figura 2.12 Estructura de una neurona del cerebro

2.5.2. Estructura de una red neuronal artificial

Evidentemente, la representación artificial de una neurona cerebral y de su funcionamiento requiere una gran simplificación. Sin embargo, como el cerebro humano, una red neuronal artificial consta de neuronas artificiales o nodos y conexiones entre ellas. Desde los nodos se transporta información a lo largo de sus conexiones a otros nodos.

Las neuronas artificiales como unidades independientes no son muy eficaces para el tratamiento de la información y se agrupan en estructuras más grandes, las redes de neuronas artificiales o redes neuronales.

La distribución de nodos dentro de la red se realiza formando niveles o capas de un número determinado de nodos cada una. A partir de su situación dentro de la red se pueden distinguir tres tipos de capas:

1. *Capas de entrada:* estas capas reciben la información desde el exterior y la forman los nodos *input*. Con ellas se identifican en los modelos estadísticos a las variables explicativas.
2. *Capas de Salida:* están formadas por los nodos *output* y envían la información hacia el exterior; identifican a las variables respuestas en los modelos estadís-

ticos.

3. *Capas ocultas*: son capas que sólo sirven para procesar información y comunicar otras capas. Sus nodos están *escondidos* y corresponden a las transformaciones de los nodos input y sus ponderaciones.

2.5.3. Los modelos de regresión como redes neuronales

Muchos de los problemas estudiados a través de una red neuronal corresponden a problemas planteados mediante modelos de regresión, en su versión más general.

Como ejemplo, en la figura 2.13 se presenta un diagrama del modelo de regresión lineal múltiple,

$$y = w_0 + \sum_{i=1}^k w_i x_i, \quad (2.2)$$

utilizando un esquema de red neuronal simple con sólo nodos input, formado por las variables explicativas x_i , $i = 1, \dots, k$ y nodos output, formado por la variable dependiente y . Los pesos sobre las flechas indican la ponderación dada a cada input x_i , formando así la suma ponderada a la que se le añade el valor w_0 para dar lugar al output y .

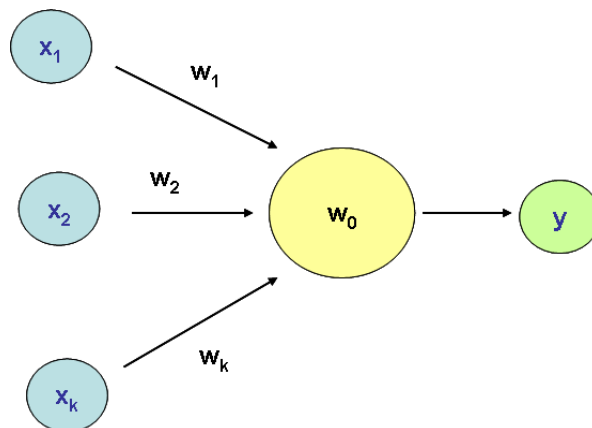


Figura 2.13 Red neuronal artificial simple

En general, hay dos aspectos importantes en el planteamiento de una red neuronal:

1. La estructura o arquitectura de la red, que se refiere a la definición de los nodos y a sus conexiones, con la que pretende representarse el problema.
2. El entrenamiento de la red a partir de los datos de entrenamiento, que permite mejorar las predicciones de la red.

Desde un punto de vista estadístico, estos dos aspectos corresponden a

1. Especificar un modelo matemático de regresión.
2. Estimar los parámetros del modelo a partir de un conjunto de datos.

La diferencia en la práctica entre ambos enfoques radica en la forma de utilizar los datos, nodos input y output en la red neuronal, para conocer los valores w_0 y \mathbf{w} . Mientras en Estadística se utilizan métodos de estimación, (máxima verosimilitud, inferencia bayesiana, estimación no paramétrica...), el enfoque de tipo *machine learning* utiliza algoritmos recursivos en los que los pesos van siendo modificados según van procesándose los datos de entrenamiento. Para más detalle puede verse Cheng y Titterington (1994).

El tipo de red neuronal que más se utiliza para resolver problemas de regresión es el de *redes Perceptrón multicapa* (*multilayer perceptron*) y generalizaciones de la *red Perceptrón simple*.

2.5.4. Redes neuronales Perceptrón.

Red Perceptrón simple

El Perceptrón es una red de alimentación directa, esto es, la información fluye desde la capa de entrada hacia la capa de salida. Fue desarrollado por F. Rosen-

blatt (1958), basándose en los modelos de neuronas biológicas de McCulloch y Pitts (1942).

La arquitectura de una red Perceptrón simple se muestra en la figura 2.13. Un conjunto de k variables input x_i binarias, generan una variable output también binaria y , a través de la expresión

$$y = f(\phi(\mathbf{x}, w)),$$

donde

$$\phi(\mathbf{x}, \mathbf{w}) = \sum_{j=1}^k w_j x_j,$$

y f es una función escalón en el valor w_0 , es decir,

$$f(u) = \begin{cases} 0 & \text{si } u < w_0 \\ 1 & \text{si } u \geq w_0 \end{cases}.$$

Formalmente,

$$y = \begin{cases} 1 & \text{si } \sum_{i=1}^k w_i x_i - w_0 \geq 0 \\ 0 & \text{si } \text{resto} \end{cases}.$$

En general, a la función f se le denomina *función de activación*.

Existen generalizaciones de esta red, utilizando otras funciones f y con variables input no necesariamente binarias. Algunos ejemplos son los siguientes:

1. $f(u) = \begin{cases} -1 & \text{si } u < 0 \\ 1 & \text{si } u \geq 0 \end{cases}$
2. $f(u) = (\text{signo}(u) + 1)/2$, que produce una respuesta binaria 0/1.
3. $f(u) = \frac{1}{1+\exp(-u)}$.
4. $f(u) = \text{tgh}(u) = \frac{e^{2u} - 1}{e^{2u} + 1}$, de forma sigmoideal y que toma valores entre -1 y 1.
5. $f(u) = u$.

En un principio las funciones de activación más utilizadas fueron las funciones escalón en un valor o umbral, pero después se extendieron a las funciones sigmoideales

(como la logística o la tangente hiperbólica) que fueron tomando mayor interés para resolver problemas de predicción.

Se puede generalizar la red Perceptrón simple al caso en el que la variable respuesta sea un vector multivariante n -dimensional. Basta tratar cada una de sus coordenadas como un output unidimensional, con ponderaciones distintas de las variables input. Así el modelo sería, para todo $j = 1, \dots, n$,

$$y_j = \begin{cases} 1 & \text{si } \sum_{i=1}^k w_{ji}x_i - w_{j0} \geq 0 \\ 0 & \text{si } \text{resto} \end{cases},$$

y el esquema de la red neuronal correspondiente puede verse en la figura 2.14.

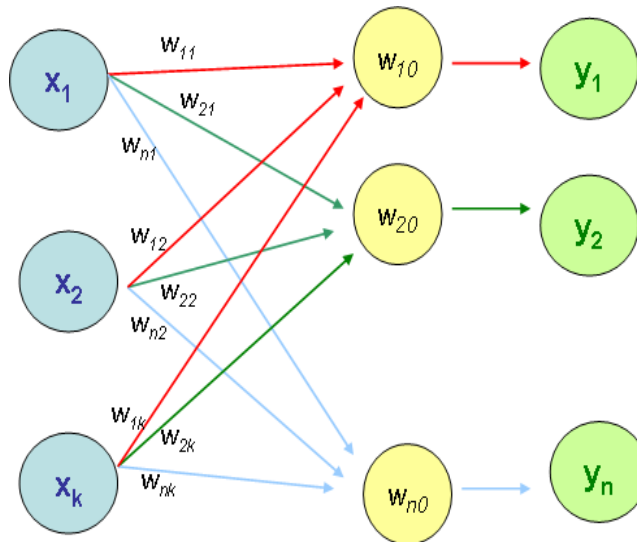


Figura 2.14 Red simple perceptrón, con output multidimensional

Red Perceptrón multicapa

Los modelos de Perceptrón multicapa, además de tener las capas de entrada y de salida, constan de al menos una capa de nodos ocultos.

Se caracterizan por que el nodo de una capa, alimenta todos los nodos de la capa siguiente, de manera que sólo puede conectarse con nodos de otras capas y, por tanto, no se permite conexiones entre nodos de la misma capa. A las redes con este tipo de conexión entre sus nodos se les denomina *redes feed-forward*.

En la figura 2.15 se observa un esquema de este tipo de redes con una sola capa de M nodos ocultos. El número de nodos de una capa oculta suele ser desconocido y es, por tanto, un parámetro más a estimar.

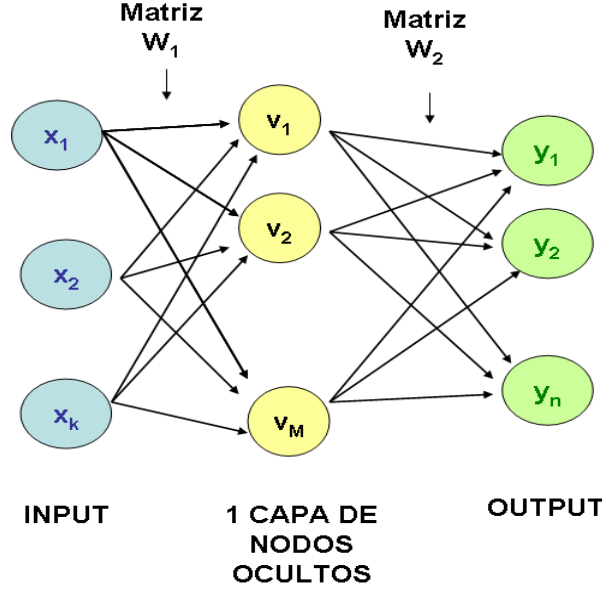


Figura 2.15 Red neuronal Feed-forward con una capa de nodos ocultos

En general, el modelo de red Perceptrón con una capa de nodos ocultos y un output n -dimensional, se puede expresar como,

$$y_j = f_2(\phi_2(\mathbf{w}_{2j}, \mathbf{v})), \quad j = 1, \dots, n$$

donde $\mathbf{W}_2 = (\mathbf{w}_{2j})_{j=1, \dots, n}$, $\mathbf{v} = (v_r)_{r=1, \dots, M}$ siendo

$$v_r = f_1(\phi_1(\mathbf{x}, \mathbf{w}_{1r})),$$

donde $\mathbf{W}_1 = (\mathbf{w}_{1r})_{r=1, \dots, M}$ y f_1 y f_2 son las funciones de activación.

Entonces, la expresión de y como una función no lineal de x equivale a una función de regresión no lineal, con parámetros \mathbf{W}_1 y \mathbf{W}_2 .

Existen varios casos de este tipo de redes, dependiendo de las funciones de activación y de las funciones ϕ_i $i = 1, 2$ que se definan.

Un ejemplo importante de redes feed-forward es aquella que tiene una capa de nodos ocultos y f_1 es la función de activación logística, es decir,

$$f_1(u) = \frac{1}{1 + \exp(-u)},$$

de modo que,

$$y_j = \beta_{j0} + \sum_{r=1}^M \beta_{jr} \frac{1}{1 + \exp\{-\gamma_{r0} - \sum_{h=1}^k \gamma_{rh} x_h\}}, \quad j = 1, \dots, n. \quad (2.3)$$

donde

$$\phi_1(\mathbf{x}, \mathbf{w}_{1r}) = \sum_{h=0}^k \gamma_{rh} x_h,$$

con $r = 1, \dots, M$, $\mathbf{w}_{1r} = (\gamma_{rh})_{h=0, \dots, k}$ y $x_0 = 1$,

$$v_r = f_1(\phi_1(\mathbf{x}, \mathbf{w}_{1r})) = \frac{1}{1 + \exp\{-\sum_{h=0}^k \gamma_{rh} x_h\}},$$

$$\phi_2(\mathbf{w}_{2j}, \mathbf{v}) = \sum_{r=0}^M \beta_{jr} v_r,$$

con $v_0 = 1$ y $\mathbf{w}_{2j} = (\beta_{jr})_{r=0, \dots, M}$ y

$$f_2(u) = u.$$

Por otro lado, Cybenko (1989) demostró que las redes neuronales de este tipo pueden aproximar uniformemente cualquier función continua utilizando un número dado M de nodos ocultos. Este resultado hace que los modelos de redes neuronales sean un buen método para resolver problemas de regresión no paramétricos.

Por último, en ocasiones, se consideran conexiones adicionales entre nodos conectando capas no contiguas. Por ejemplo, en el modelo (2.3), se pueden considerar también como inputs, tanto la combinación lineal de las funciones logísticas (nodos ocultos) como una combinación lineal de los nodos input, es decir,

$$y_{ij} = \beta_{j0} + \lambda^T \mathbf{x} + \sum_{k=1}^M \beta_{jk} \frac{1}{1 + \exp\{-\gamma_{k0} - \sum_{h=1}^r \gamma_{kh} x_{ih}\}}.$$

Estos modelos de redes neuronales con una capa de redes ocultas van a servir de base para construir los modelos de regresión no lineales que se utilizarán en el capítulo 4.

2.6. Escalamiento multidimensional: análisis INDSCAL

Los resultados desarrollados en esta tesis se aplican a conjuntos de puntos representados en \mathbb{R}^2 . En concreto, se representan conjuntos de genes en situaciones diferentes con el fin de proceder al estudio de su emparejamiento y su solapamiento. Con el objetivo de conseguir la representación de los genes en el espacio \mathbb{R}^2 , se parte de sus expresiones y se calculan las distancias euclídeas entre ellos, definiendo una matriz de distancias. La técnica INDSCAL permite la representación de estos genes proyectándolos, normalmente, en un espacio de dos dimensiones.

El origen del escalamiento multidimensional tuvo lugar en el área de la Psicología, pero actualmente se está extendiendo a otros campos de aplicación. En términos generales, el escalamiento multidimensional es una técnica estadística multivariante que, tomando como datos una matriz de proximidad o similitud entre objetos, proporciona una proyección de éstos en un espacio de dimensión r (generalmente dos o tres), donde las distancias entre ellos se relaciona mediante una función lineal con las similitudes (escalamiento métrico).

Cuando existen varias matrices de similitud entre objetos (por ejemplo, dadas por distintos individuos), es interesante poder ponderar de diferente manera las dimensiones de la representación de los objetos, dependiendo del individuo. Esto es lo que consigue el análisis INDSCAL, tomando como entrada varias matrices de proximidad, una para cada uno de los individuos. El modelo supone que hay un conjunto de r dimensiones comunes a todos los objetos, pero que las distancias entre ellos en este espacio cambian de un individuo a otro, de acuerdo a la importancia o peso que cada uno dé a cada dimensión. Así, denotando como $s_{jk}^{(i)}$ la similitud entre

el objeto j y el k dada por el individuo i , el modelo asume que

$$s_{jk}^{(i)} = L(d_{jk}^{(i)}),$$

donde L es una función lineal (con pendiente negativa) y $d_{jk}^{(i)}$ es la distancia euclídea entre los objetos ponderada por distintos pesos, es decir,

$$d_{jk}^{(i)} = \sqrt{\sum_{t=1}^r w_{it}(x_{jt} - x_{kt})^2},$$

donde x_{jt} es la coordenada del j -ésimo objeto en la dimensión t . Se puede observar que $d_{jk}^{(i)}$ puede interpretarse como la distancia euclídea ordinaria entre los puntos

$$y_{jt}^{(i)} = \sqrt{w_{it}}x_{jt},$$

es decir, la configuración es contraída o expandida dependiendo de los valores de las raíces cuadradas de los pesos.

El resultado del análisis proporciona, por una parte, las coordenadas de todos los objetos sin diferenciar por individuo en el denominado espacio global de objetos o espacio de estímulos, (frecuentemente de dimensión $r = 2$) y, por otra, los pesos o ponderaciones que cada individuo asigna a cada dimensión.

Las raíces cuadradas de las ponderaciones de cada individuo, se representan como vectores en un espacio de dimensión dos denominado espacio de sujetos. Cuanto menor sea el ángulo entre el vector de un individuo y una dimensión dada, mayor es la ponderación o la importancia que está dando a esa dimensión. Por otra parte el ángulo formado entre los vectores de dos individuos, indicará el grado de similitud entre ambos.

Multiplicando las coordenadas de cada objeto en el espacio global de objetos por las raíces cuadradas de sus ponderaciones se consigue la representación de los objetos por cada sujeto o individuo en el mismo espacio r -dimensional. (véase por ejemplo Borg y Groenen, 1997).

Capítulo 3

Emparejamiento de más de dos configuraciones no etiquetadas mediante transformaciones lineales

3.1. Introducción

En este capítulo se estudiará el problema general del emparejamiento de $M > 2$ configuraciones de puntos, como generalización del modelo Green y Mardia (2006), y se irán desarrollando diferentes situaciones dependiendo del tipo de transformaciones que puedan existir entre las configuraciones consideradas.

En primer lugar se realizará una exposición del problema a resolver y se presentará el desarrollo del modelo matemático en su versión más general, que incluirá la definición del Proceso de Poisson que da lugar a la configuración de referencia, y las matrices de afinidad y de emparejamientos de distintos órdenes. A continuación, se presentarán distintos modelos variando el tipo de transformaciones, empezando por las transformaciones lineales generales que modelizan deformaciones entre configuraciones, para pasar a tratar el modelo con transformaciones de similitud (rotaciones, traslaciones y escalamientos) más utilizadas en el análisis Procrustes (Dryden y Mardia, 1998) al mantener constantes los cocientes entre distancias entre

puntos y, por lo tanto, mantener la forma de las configuraciones.

Al final del capítulo se presentan aplicaciones de estas situaciones, inicialmente con datos simulados para comprobar la validez de los resultados teóricos, y posteriormente, y como aplicación real en Bioinformática, con datos provenientes de experimentos con microarrays.

3.2. Modelización del problema

Partimos de M configuraciones con $M > 2$ situadas en el espacio \mathbb{R}^d donde $d \geq 2$. Cada una de ellas está formada por n_i puntos ($i = 1, \dots, M$), de modo que,

$$\mathbf{x}_i = \{x_{ij}, j = 1, \dots, n_i\} \quad x_{ij} \in \mathbb{R}^d \quad i = 1, \dots, M.$$

El problema es determinar de qué manera están emparejadas estas configuraciones, es decir, qué puntos de cada configuración están emparejados con otros puntos de otras configuraciones teniendo en cuenta que existen transformaciones geométricas entre los sistemas de referencia donde están representados o, equivalentemente, entre las configuraciones.

Al tratarse de más de dos configuraciones, los emparejamientos pueden ser de distintos órdenes. Si fueran sólo dos, las posibilidades de emparejamiento serían simples: cada punto de una configuración,

- i) o está emparejado con un punto de la otra configuración
- ii) o no está emparejado con ninguno.

Si fueran tres configuraciones, las posibilidades de emparejamiento se amplían: cada punto de una configuración,

- i) o no está emparejado con ningún punto de ninguna de las otras dos configuraciones

- ii) o tiene emparejamiento con un punto de una de las dos configuraciones
- iii) o está emparejado con dos puntos, uno de cada una de las otras dos configuraciones.

Por tanto, pueden presentarse emparejamientos dobles o emparejamientos triples. Así, la situación se generaliza pudiéndose presentar, emparejamientos dobles, triples, cuádruples,..., hasta emparejamientos de orden M .

Suponemos que estas configuraciones son perturbaciones aleatorias de un conjunto de puntos fijos y desconocidos $\{\mu_l\} \in \mathbb{R}^d$, (la configuración de referencia), y que cada μ_l sólo puede generar a lo sumo un punto de una misma configuración pero más de un punto de distintas configuraciones. Se desconoce, qué puntos de cada configuración se corresponden con cada uno de estos μ_l , pero aquellos puntos de configuraciones distintas que son perturbaciones aleatorias de un mismo punto μ_l estarán emparejados.

La correspondencia entre puntos μ_l y puntos de las configuraciones podría representarse a través de una función entre ellos, de manera que a cada x_{ij} , le correspondería el μ_l que lo genera, verificando que si $j \neq j'$ el punto μ_l que genera a x_{ij} es distinto al que genera $x_{ij'}$.

Otra opción es definir estas correspondencias entre los subíndices de $\{\mu_l\}$ y los de los datos $\{x_{ij}\}$ $i = 1, \dots, M$, $j = 1, \dots, n_i$ indicando qué punto μ_l está asociado con qué puntos de los \mathbf{x}_i a través de la matriz de índices $\{\xi_{ij}\}$ $i = 1, \dots, M$, $j = 1, \dots, n_i$. Así ξ_{ij} representa el subíndice de μ_l que genera el punto j de la configuración i , es decir, el punto x_{ij} viene generado a partir de $\mu_{\xi_{ij}}$.

Además, existen unas transformaciones geométricas desconocidas entre los sistemas de referencia donde se representan las M configuraciones, que también formarán parte del modelo.

Sin pérdida de generalidad, consideramos la configuración \mathbf{x}_1 en el sistema de

coordenadas de referencia donde se encuentra $\{\mu_l\}$. Así, consideramos que los puntos de \mathbf{x}_1 se generan a partir de los $\{\mu_l\}$ más un error aleatorio y la transformación geométrica entre el sistema de coordenadas de \mathbf{x}_1 y el de la configuración \mathbf{x}_i la denotaremos por ϕ_{i-1} $i = 2, \dots, M$.

Bajo estas hipótesis, el modelo matemático que refleja esta situación va a venir dado por,

$$\begin{aligned}
 x_{1j} &= \mu_{\xi_{1j}} + \varepsilon_{1j} & j &= 1, \dots, n_1 \\
 \phi_1(x_{2j}) &= \mu_{\xi_{2j}} + \varepsilon_{2j} & j &= 1, \dots, n_2 \\
 &\vdots & &\vdots \\
 \phi_{r-1}(x_{rj}) &= \mu_{\xi_{rj}} + \varepsilon_{rj} & j &= 1, \dots, n_r \\
 &\vdots & &\vdots \\
 \phi_{M-1}(x_{Mj}) &= \mu_{\xi_{Mj}} + \varepsilon_{Mj} & j &= 1, \dots, n_M
 \end{aligned} \tag{3.1}$$

donde para todo $i = 1, \dots, M$ y para todo $j = 1, \dots, n_i$, ε_{ij} tiene como función de densidad f_i . Además, se supone que cada μ_l genera a lo sumo, un punto de cada configuración, es decir, cada punto de una configuración tiene a lo sumo un emparejamiento, por lo que, fijado $i = 1, \dots, M$, $\xi_{i1} \neq \xi_{i2} \neq \dots \neq \xi_{in_i}$. Por último, todos los $\{\varepsilon_{ij}\}$ son independientes entre sí e independientes de los $\{\mu_l\}$.

Una vez planteado el modelo matemático que representa la situación a estudiar, vamos a ir definiendo distintos elementos del mismo, comenzando por el Proceso de Poisson que da lugar a los puntos fijos $\{\mu_l\}$ de la configuración de referencia.

3.3. El Proceso de Poisson espacial de la configuración de referencia

Supongamos que el conjunto de puntos fijos $\{\mu_l\}$ forman un proceso de Poisson homogéneo de tasa λ sobre una región $V \subset \mathbb{R}^d$ de volumen v , y que partimos de N observaciones de este proceso en dicha región. Suponemos además que cada uno de estos μ_l puede generar de forma independiente, ningún punto de ninguna configuración, un único punto de una configuración, dos puntos de dos configuraciones

distintas (es decir, un emparejamiento doble), tres puntos de tres configuraciones distintas, es decir, un emparejamiento triple, y así hasta M puntos, uno de cada configuración. Así, los $\{\mu_l\}$ quedan clasificados en $M + 1$ clases distintas e independientes, que denotaremos por C_0, C_1, \dots, C_{M-1} , y C_M , respectivamente, dependiendo del tipo de emparejamiento que generan.

Para calcular la probabilidad de pertenecer a una clase o a otra, consideramos unas tasas de emparejamientos a priori ρ_s , $s = 1, \dots, M$, entendidas como medidas de las tendencias a generar emparejamientos de distintos órdenes. Estas son independientes de cuáles sean las configuraciones involucradas en los emparejamientos, aunque pueden ser distintas dependiendo del orden del emparejamiento. Por ejemplo, se podría pensar que las tasas de emparejamientos de órdenes superiores son menores que las de emparejamientos de órdenes inferiores. Todo esto queda resumido en las siguientes suposiciones sobre los puntos del proceso de Poisson espacial:

1. Cada μ_l tiene la misma probabilidad p de generar un punto de una configuración \mathbf{x}_i o \mathbf{x}_j .
2. Las tasas de emparejamientos a priori son distintas dependiendo del orden de emparejamiento.
3. La tasa de emparejamiento de cierto orden a priori es independiente de cuáles sean las configuraciones a las que pertenezcan los puntos emparejados.

Bajo estas hipótesis, un punto μ_l pertenecerá a la clase C_i , $i = 1, \dots, M$ con probabilidad $\rho_i p^i$ y a la clase C_0 con probabilidad $1 - \sum_{j=1}^M \rho_j p^j$ con $\rho_1 = 1$.

Una vez definido el Proceso de Poisson que genera la configuración de referencia, vamos a definir las matrices de afinidad y de emparejamiento que permitirán identificar los emparejamientos entre configuraciones.

3.4. Matrices de afinidad M y matrices de emparejamiento S

Con el objetivo de identificar cuáles son los puntos que están emparejados definimos, en primer lugar, las *matrices de afinidad* con las que se recogerá el número de veces que cada configuración está emparejada con las otras configuraciones y, en segundo lugar, las *matrices de emparejamiento* con las que, una vez conocidas qué configuraciones son las que tienen al menos un punto emparejado, se identificarán cuáles son estos puntos.

La definición de los emparejamientos a través de estos dos tipos de matrices, permite modelizar situaciones donde se tiene la información parcial de cuántos emparejamientos hay entre configuraciones, lo que redundaría en un conocimiento de las matrices de afinidad, pero se desconoce cuáles son exactamente los puntos involucrados en los emparejamientos.

3.4.1. Matriz de afinidad y matriz de emparejamientos de orden 2

Definición 25 *Se define la matriz de afinidad de orden 2, $M^{(2)}$ de dimensión $M \times M$, como aquella matriz cuyos elementos son:*

$$\{M_{i_1 i_2}^{(2)}\} = \text{número de emparejamientos dobles que hay entre las configuraciones } \mathbf{x}_{i_1} \text{ y } \mathbf{x}_{i_2}.$$

Esta matriz será simétrica con ceros en la diagonal. Además, si denotamos con L_i , $i = 2, \dots, M$, el número total de emparejamientos de orden i que hay entre todas las configuraciones,

$$\sum_{i_1=1}^M \sum_{i_2 > i_1}^M M_{i_1 i_2}^{(2)} = L_2.$$

Una vez que la matriz de afinidad de orden 2 identifica qué configuraciones tienen algún emparejamiento doble y cuántos son, es necesario identificar los puntos que

están emparejados.

Definición 26 Para todo $i_1 < i_2$, ($i_1, i_2 \in \{1, \dots, M\}$), tales que $M_{i_1 i_2}^{(2)} \neq 0$, se define la matriz de emparejamientos de orden 2 de las configuraciones \mathbf{x}_{i_1} y \mathbf{x}_{i_2} y se denota por $\mathbf{S}^{(i_1, i_2)}$, como aquella matriz de dimensiones $n_{i_1} \times n_{i_2}$, cuyos elementos son:

$$S_{j_1 j_2}^{(i_1, i_2)} = \begin{cases} 1 & \text{si } \xi_{i_1 j_1} = \xi_{i_2 j_2} \\ 0 & \text{resto} \end{cases}.$$

Por tanto, los superíndices van a identificar las configuraciones que son afines (por existir emparejamientos entre ellas), y los subíndices los puntos de esas configuraciones que están emparejados. Así, el número total de emparejamientos dobles entre las configuraciones afines \mathbf{x}_{i_1} y \mathbf{x}_{i_2} se podrá expresar como:

$$M_{i_1 i_2}^{(2)} = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} S_{j_1 j_2}^{(i_1, i_2)}.$$

3.4.2. Matriz de afinidad y matriz de emparejamientos de orden 3

Definición 27 Se define la matriz de afinidad de orden 3, $\mathbf{M}^{(3)}$, de dimensión $M \times M \times M$, como aquella array cuyos elementos son:

$$\{M_{i_1 i_2 i_3}^{(3)}\} = \text{número de emparejamientos triples que hay entre las configuraciones } \mathbf{x}_{i_1}, \mathbf{x}_{i_2} \text{ y } \mathbf{x}_{i_3}.$$

Esta matriz será una matriz tridimensional con ceros en todas las coordenadas tales que $x = y$, $x = z$ ó $y = z$. Además,

$$\sum_{i_1=1}^M \sum_{i_2 > i_1}^M \sum_{i_3 > i_2}^M M_{i_1 i_2 i_3}^{(3)} = L_3.$$

Una vez que la matriz de afinidad de orden tres identifica qué configuraciones tienen algún emparejamiento triple, definimos la matriz de emparejamientos de orden 3.

Definición 28 Para todo $i_1 < i_2 < i_3$, $(i_1, i_2, i_3 \in \{1, \dots, M\})$ tales que $M_{i_1 i_2 i_3}^{(3)} \neq 0$, se define la matriz de emparejamientos de orden 3 de las configuraciones \mathbf{x}_{i_1} , \mathbf{x}_{i_2} y \mathbf{x}_{i_3} y se denota por $\mathbf{S}^{(i_1, i_2, i_3)}$, como aquel array de dimensiones $n_{i_1} \times n_{i_2} \times n_{i_3}$ cuyos elementos son:

$$S_{j_1 j_2 j_3}^{(i_1, i_2, i_3)} = \begin{cases} 1 & \text{si } \xi_{i_1 j_1} = \xi_{i_2 j_2} = \xi_{i_3 j_3} \\ 0 & \text{resto} \end{cases}.$$

Así, el número total de emparejamientos triples entre las configuraciones \mathbf{x}_{i_1} , \mathbf{x}_{i_2} y \mathbf{x}_{i_3} se podrá expresar como:

$$M_{i_1 i_2 i_3}^{(3)} = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \sum_{j_3=1}^{n_3} S_{j_1 j_2 j_3}^{(i_1, i_2, i_3)}.$$

3.4.3. Matriz de afinidad y matriz de emparejamientos de orden k

Definición 29 En general, se define la matriz de afinidad de orden k , $\mathbf{M}^{(k)}$, de dimensión $M \times \binom{k}{\dots} \times M$, como aquella array cuyos elementos son:

$$\{M_{i_1 \dots i_k}^{(k)}\} = \text{número de emparejamientos de orden } k \text{ que hay} \\ \text{entre las configuraciones } \mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_k}.$$

Esta matriz será una matriz k -dimensional con ceros en todas las coordenadas tales que $x_{i_r} = x_{i_s}$. Además,

$$\sum_{i_1=1}^M \sum_{i_2 > i_1}^M \dots \sum_{i_k > i_{k-1}}^M M_{i_1 \dots i_k}^{(k)} = L_k.$$

Definición 30 Para todo $i_1 < i_2 < \dots < i_{k-1} < i_k$ ($i_1, \dots, i_k \in \{1, \dots, M\}$) tales que $M_{i_1 \dots i_k}^{(k)} \neq 0$, se define la matriz de emparejamientos de orden k de las configuraciones $\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_k}$ y se denota por $\mathbf{S}^{(i_1, \dots, i_k)}$, como aquel array de dimensiones $n_{i_1} \times \dots \times n_{i_k}$ cuyos elementos son:

$$S_{j_1 \dots j_k}^{(i_1, \dots, i_k)} = \begin{cases} 1 & \text{si } \xi_{i_1 j_1} = \xi_{i_2 j_2} = \dots = \xi_{i_k j_k} \\ 0 & \text{resto} \end{cases}.$$

Así, el número total de emparejamientos de orden k entre los k configuraciones se podrá expresar como:

$$M_{i_1 \dots i_k}^{(k)} = \sum_{j_1=1}^{n_1} \dots \sum_{j_k=1}^{n_k} S_{j_1 \dots j_k}^{(i_1, \dots, i_k)}.$$

Se puede observar que en el caso de $k = M$, $M_{1 \dots M}^{(M)} = L_M$.

Ejemplo 31 Supongamos $M = 4$ configuraciones, cada una de ellas con $n_i = 5$ puntos. Supongamos que $L_2 = 3$ y $L_3 = 2$. Esto indica que hay tres emparejamientos dobles, dos triples y, por tanto, un total de ocho puntos que no están emparejados. Supongamos que, los emparejamientos dobles son $(x_{11}, x_{21}), (x_{24}, x_{32})$ y (x_{23}, x_{31}) , y que los emparejamientos triples son (x_{12}, x_{22}, x_{41}) y (x_{14}, x_{33}, x_{44}) .

Entonces la matriz de afinidad de orden 2 será:

$$\mathbf{M}^{(2)} = \begin{array}{ccccc} & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \\ \mathbf{x}_1 & 0 & 1 & 0 & 0 \\ \mathbf{x}_2 & 1 & 0 & 2 & 0 \\ \mathbf{x}_3 & 0 & 2 & 0 & 0 \\ \mathbf{x}_4 & 0 & 0 & 0 & 0 \end{array},$$

es decir, hay un emparejamiento doble entre las configuraciones \mathbf{x}_1 y \mathbf{x}_2 y dos entre las configuraciones \mathbf{x}_2 y \mathbf{x}_3 .

Las matrices de emparejamientos dobles serán,

$$\mathbf{S}^{(1,2)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{S}^{(2,3)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

es decir, están emparejados los puntos 1 y 1 de las configuraciones \mathbf{x}_1 y \mathbf{x}_2 , y los puntos 3 y 1 y los puntos 4 y 2 de las configuraciones \mathbf{x}_2 y \mathbf{x}_3 .

Sobre los emparejamientos de tercer orden, al tratarse $\mathbf{M}^{(3)}$ de una matriz en tres dimensiones, para cada $i = 1, \dots, 4$, calculamos las 4 matrices de dimensión dos $(\mathbf{M}_i^{(3)})_{i=1,\dots,4}$, donde $(\mathbf{M}_i^{(3)})$ recogerá los emparejamientos de orden 3 en los que está involucrada la configuración \mathbf{x}_i con el resto de las configuraciones (corresponderían a las 4 "slices", fijada la primera coordenada de la matriz tridimensional).

Así,

$$\begin{aligned} \mathbf{M}_1^{(3)} &= \begin{array}{cc} & \begin{matrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \end{matrix} \\ \begin{matrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{array}, \mathbf{M}_2^{(3)} = \begin{array}{cc} & \begin{matrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \end{matrix} \\ \begin{matrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \\ \mathbf{M}_3^{(3)} &= \begin{array}{cc} & \begin{matrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \end{matrix} \\ \begin{matrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array}, \mathbf{M}_4^{(3)} = \begin{array}{cc} & \begin{matrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \end{matrix} \\ \begin{matrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \end{matrix} & \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \end{aligned}$$

es decir, hay un emparejamiento triple entre las configuraciones $\mathbf{x}_1, \mathbf{x}_2$ y \mathbf{x}_4 y otro emparejamiento triple entre las configuraciones $\mathbf{x}_1, \mathbf{x}_3$ y \mathbf{x}_4 .

Para calcular las matrices de emparejamientos $\mathbf{S}^{(1,2,4)}$ y $\mathbf{S}^{(1,3,4)}$, que vuelven a ser matrices tridimensionales, procedemos de la misma manera (en este caso, sólo se muestran aquellas matrices con algún elemento distinto de cero).

$$\mathbf{S}_2^{(1,2,4)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{S}_4^{(1,3,4)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Por tanto, hemos definido unos nuevos parámetros, las matrices de afinidad y de emparejamientos, con los que se describen los emparejamientos existentes. Además, las funciones ϕ_i de (3.1) representan las transformaciones geométricas que relacionan a las M configuraciones. Todos estos parámetros, junto con los que están involucrados en las distribuciones f_i de los ε_{ij} de (3.1), forman el total de parámetros de nuestro modelo.

3.4.4. Distribuciones a priori de las matrices de afinidad y de emparejamientos

Distribución a priori de las matrices de afinidad

Vamos a definir la distribución a priori de las matrices de afinidad a partir de su distribución condicionada por el número de emparejamientos que hay de cada orden.

Supongamos que la distribución de la matriz de afinidad de orden k , sabiendo que hay L_k emparejamientos de orden k , es la distribución uniforme. Esto equivale a suponer que conocidos cuántos emparejamientos de orden k hay, todas las posibles combinaciones de L_k emparejamientos de orden k son igual de probables.

Por ejemplo, si hay $L_2 = 4$ emparejamientos dobles entre un grupo de $M = 3$ configuraciones, hay 15 maneras posibles de presentarse estos emparejamientos, todas ellas con la misma probabilidad:

	Distribución de los $L_2 = 4$ emparejamientos													
$\mathbf{x}_1, \mathbf{x}_2$	4	3	3	2	2	2	1	1	1	1	0	0	0	0
$\mathbf{x}_1, \mathbf{x}_3$	0	1	0	1	2	0	3	0	1	2	4	0	3	1
$\mathbf{x}_2, \mathbf{x}_3$	0	0	1	1	0	2	0	3	2	1	0	4	1	3

Por lo tanto, en general hay que calcular cuántas matrices de afinidad de orden k puede haber, sabiendo que hay L_k emparejamientos de orden k .

Suponiendo n_i suficientemente grande, esto equivale a distribuir L_k emparejamientos en $\binom{M}{k}$ posiciones (las posiciones representan todos los posibles emparejamientos de orden k que puede haber con las M configuraciones). Así, el número de matrices de afinidad de orden k que puede haber se calcula como combinaciones con repetición de $\binom{M}{k}$ configuraciones tomados de L_k en L_k .

En el ejemplo $\binom{M}{k} = \binom{3}{2} = 3$, y el número de combinaciones con repetición de $M = 3$ configuraciones tomados de $L_2 = 4$ en 4, es $\binom{\binom{M}{k} + L_k - 1}{L_k} = \binom{3+4-1}{4} = \binom{6}{4} = 15$. Por tanto,

$$p(\mathbf{M}^{(k)} \mid L_k) = \frac{1}{\binom{\binom{M}{k} + L_k - 1}{L_k}}.$$

Así, asumiendo independencia para configuraciones con una cantidad de puntos suficientemente grande, la distribución conjunta de todas ellas condicionada por los números de emparejamientos, será de la forma:

$$p(\mathbf{M}^{(2)}, \mathbf{M}^{(3)}, \dots, \mathbf{M}^{(M)} \mid L_2, L_3, \dots, L_M) = \prod_{k=1}^M \frac{1}{\binom{\binom{M}{k} + L_k - 1}{L_k}}. \quad (3.2)$$

Para obtener la distribución a priori de estas matrices de afinidad, bastará multiplicar (3.2) por la distribución conjunta a priori del número de emparejamientos.

Supongamos que N, n_1, \dots, n_M son conocidos y sea $n = \sum_{i=1}^M n_i$. Entonces se observarán L_2 emparejamientos dobles, L_3 emparejamientos triples, ..., L_{M-1} emparejamientos de $M-1$ configuraciones y L_M emparejamientos de las M configuraciones, siempre y cuando el número de μ_l de cada clase C_k ($k = 0, \dots, M$),

definidas en (3.3) sean, respectivamente $N - n + L_2 + 2L_3 + \dots + (M - 1)L_M$, $n - 2L_2 - 3L_3 - \dots - ML_M$, L_2, \dots, L_M .

Bajo las suposiciones presentadas en (3.3), las $M+1$ variables aleatorias definidas como número de puntos fijos μ_l de la clase C_k observados en el volumen v ($k = 0, \dots, M$) son independientes y con distribución de Poisson con tasas $\lambda v(1 - \sum_{j=1}^M \rho_j p^j)$, $\lambda v p$, $\lambda v \rho_2 p^2, \dots, \lambda v \rho_M p^M$, respectivamente.

Clase de μ_l	Frecuencia	Tasa
C_0	$N - n + L_2 + 2L_3 + \dots + (M - 1)L_M$	$\lambda v(1 - \sum_{j=1}^M \rho_j p^j)$
C_1	$n - 2L_2 - 3L_3 - \dots - ML_M$	$\lambda v p$
C_2	L_2	$\lambda v \rho_2 p^2$
\dots	\dots	\dots
C_{M-1}	L_{M-1}	$\lambda v \rho_{M-1} p^{M-1}$
C_M	L_M	$\lambda v \rho_M p^M$

Así, la distribución a priori conjunta de L_2, L_3, \dots, L_M condicionada por n_1, \dots, n_M será:

$$\begin{aligned}
 p(L_2, L_3, \dots, L_M) &\propto \frac{e^{-\lambda v p} (\lambda v p)^{n - \sum_{i=2}^M i L_i}}{(n - \sum_{i=2}^M i L_i)!} \cdot \frac{e^{-\lambda v \rho_2 p^2} (\lambda v \rho_2 p^2)^{L_2}}{L_2!} \cdot \dots \cdot \\
 &\frac{e^{-\lambda v \rho_M p^M} (\lambda v \rho_M p^M)^{L_M}}{L_M!} = \\
 &= \frac{e^{-\lambda v (\sum_{i=1}^M \rho_i p^i)} (\lambda v)^{n - \sum_{i=2}^M (i-1) L_i} p^n \rho_2^{L_2} \rho_3^{L_3} \dots \rho_M^{L_M}}{(n - \sum_{i=2}^M i L_i)! L_2! L_3! \dots L_M!},
 \end{aligned}$$

por lo que, considerando constante todo lo que no depende de L_2, L_3, \dots, L_M , se tiene que:

$$p(L_2, L_3, \dots, L_M) \propto \frac{(\frac{\rho_2}{\lambda v})^{L_2} \dots (\frac{\rho_M}{\lambda v})^{L_M}}{(\lambda v)^{L_3 + 2L_4 + \dots + (M-2)L_M} (n - \sum_{i=2}^M i L_i)! L_2! L_3! \dots L_M!}, \quad (3.3)$$

donde $\sum_{i=2}^M i L_i \leq n$.

Por lo tanto, multiplicando (3.2) y (3.3) se tiene que,

$$\begin{aligned}
 p(\mathbf{M}^{(2)}, \mathbf{M}^{(3)}, \dots, \mathbf{M}^{(M)}) &\propto p(\mathbf{M}^{(2)}, \mathbf{M}^{(3)}, \dots, \mathbf{M}^{(M)} \mid L_2, L_3, \dots, L_M) p(L_2, L_3, \dots, L_M) \\
 &\propto \prod_{k=1}^M \frac{1}{\binom{M}{k} + L_k - 1} \frac{(\frac{\rho_2}{\lambda v})^{L_2} \dots (\frac{\rho_M}{\lambda v})^{L_M}}{(\lambda v)^{\sum_{k=3}^M (k-2) L_k} (n - \sum_{i=2}^M i L_i)! L_2! L_3! \dots L_M!}.
 \end{aligned}$$

Operando y considerando constante todo aquello que no dependa de $\mathbf{M}^{(i)}$, la distribución a priori conjunta de las matrices de afinidad tiene como expresión:

$$p(\mathbf{M}^{(2)}, \mathbf{M}^{(3)}, \dots, \mathbf{M}^{(M)}) \propto \frac{(\frac{\rho_2}{\lambda v})^{L_2} \dots (\frac{\rho_M}{\lambda v})^{L_M}}{(\lambda v)^{\sum_{k=3}^M (k-2)L_k} (n - \sum_{i=2}^M iL_i)! \prod_{k=2}^M [\binom{M}{k} + L_k - 1]!}. \quad (3.4)$$

Distribución a priori de las matrices de emparejamiento

Consideramos ahora la distribución a priori de las matrices de emparejamiento. Supongamos, por ejemplo, conocida la matriz de afinidad de orden dos $\mathbf{M}^{(2)}$. Como por hipótesis todos los emparejamientos de orden 2 son igualmente probables (es decir, no dependen de las configuraciones que se emparejen) podemos asumir una distribución uniforme sobre todas las matrices de emparejamientos de orden 2. Así, debemos calcular, para cada pareja (i_1, i_2) tales que $\mathbf{M}_{i_1, i_2}^{(2)} \neq 0$, el número de matrices de emparejamientos $\mathbf{S}^{(i_1, i_2)}$ que se pueden construir.

Conocido $\mathbf{M}_{i_1, i_2}^{(2)} \neq 0$ primero seleccionamos los puntos de la configuración \mathbf{x}_{i_1} que van a estar emparejados con los de la configuración \mathbf{x}_{i_2} (en total, $\binom{n_{i_1}}{\mathbf{M}_{i_1, i_2}^{(2)}}$). A continuación hacemos lo mismo con los puntos de la configuración \mathbf{x}_{i_2} (en total, $\binom{n_{i_2}}{\mathbf{M}_{i_1, i_2}^{(2)}}$), y después, fijando los puntos de una de las configuraciones y permutando los puntos de la otra configuración se generan todos los posibles emparejamientos, (en total $\mathbf{M}_{i_1, i_2}^{(2)}!$).

Por tanto, para cada pareja (i_1, i_2) tales que $\mathbf{M}_{i_1, i_2}^{(2)} \neq 0$, la distribución de la matriz de emparejamientos $\mathbf{S}^{(i_1, i_2)}$, dada la matriz de afinidad de orden 2 será:

$$p(\mathbf{S}^{(i_1, i_2)} | \mathbf{M}^{(2)}) = \frac{1}{\left[\binom{n_{i_1}}{\mathbf{M}_{i_1, i_2}^{(2)}} \binom{n_{i_2}}{\mathbf{M}_{i_1, i_2}^{(2)}} \mathbf{M}_{i_1, i_2}^{(2)}! \right]}.$$

Asumiendo independencia, la distribución conjunta de todas las matrices de em-

parejamientos de orden dos, dada la matriz de afinidad de orden dos, será:

$$p(\mathbf{S}^{(1,2)}, \dots, \mathbf{S}^{(M-1,M)} | \mathbf{M}^{(2)}) = \frac{1}{\prod_{\{(i_1, i_2) | \mathbf{M}_{i_1, i_2}^{(2)} \neq 0\}} \binom{n_{i_1}}{\mathbf{M}_{i_1, i_2}^{(2)}} \binom{n_{i_2}}{\mathbf{M}_{i_1, i_2}^{(2)}} \mathbf{M}_{i_1, i_2}^{(2)}!}.$$

Razonando de la misma forma, se puede obtener la distribución conjunta de todas las matrices de emparejamientos de orden 3, dada la matriz de afinidad de orden 3. Para cada terna (i_1, i_2, i_3) tales que $\mathbf{M}_{i_1, i_2, i_3}^{(3)} \neq 0$, seleccionamos los puntos de la configuración \mathbf{x}_{i_1} que van a estar emparejados (en total, $\binom{n_{i_1}}{\mathbf{M}_{i_1, i_2, i_3}^{(3)}}$). Después hacemos lo mismo con los puntos de la configuración \mathbf{x}_{i_2} (en total, $\binom{n_{i_2}}{\mathbf{M}_{i_1, i_2, i_3}^{(3)}}$) y con los de la tercera configuración \mathbf{x}_{i_3} (en total, $\binom{n_{i_3}}{\mathbf{M}_{i_1, i_2, i_3}^{(3)}}$) y por último permutamos los puntos de dos configuraciones para generar todos los posibles emparejamientos $\left(\left(\mathbf{M}_{i_1, i_2, i_3}^{(3)}!\right)^2\right)$.

Por tanto, de forma análoga al caso anterior, la distribución conjunta de las matrices de emparejamientos de orden tres, dada la matriz de afinidad de orden tres será:

$$p(\mathbf{S}^{(1,2,3)}, \dots, \mathbf{S}^{(M-2, M-1, M)} | \mathbf{M}^{(3)}) = \frac{1}{\prod_{\{(i_1, i_2, i_3) | \mathbf{M}_{i_1, i_2, i_3}^{(3)} \neq 0\}} \binom{n_{i_1}}{\mathbf{M}_{i_1, i_2, i_3}^{(3)}} \binom{n_{i_2}}{\mathbf{M}_{i_1, i_2, i_3}^{(3)}} \binom{n_{i_3}}{\mathbf{M}_{i_1, i_2, i_3}^{(3)}} \left(\mathbf{M}_{i_1, i_2, i_3}^{(3)}!\right)^2}.$$

En general, para todo $k = 2, \dots, M$, la distribución conjunta de las matrices de emparejamientos de orden k , dada la matriz de afinidad de orden k , será:

$$p(\mathbf{S}^{(1, \dots, k)}, \dots, \mathbf{S}^{(M-(k-1), \dots, M)} | \mathbf{M}^{(k)}) = \frac{1}{\prod_{\substack{\{(i_1, \dots, i_k) | \\ \mathbf{M}_{i_1, \dots, i_k}^{(k)} \neq 0\}}} \binom{n_{i_1}}{\mathbf{M}_{i_1, \dots, i_k}^{(k)}} \cdots \binom{n_{i_k}}{\mathbf{M}_{i_1, \dots, i_k}^{(k)}} \left(\mathbf{M}_{i_1, \dots, i_k}^{(k)}!\right)^{k-1}}. \quad (3.5)$$

Por tanto, el producto de todas ellas, dará lugar a la distribución a priori conjunta de todas las matrices de emparejamientos, dadas las matrices de afinidad.

A continuación vamos a estudiar distintas situaciones, dependiendo de la transformación existente entre los sistemas de coordenadas donde se representan las con-

figuraciones. Partiremos del caso más general, donde se supone que las transformaciones entre configuraciones son transformaciones lineales generales, para pasar al caso de transformaciones de similitud (rotaciones, traslaciones y escalamientos).

3.5. Modelización con transformaciones lineales generales

Una transformación lineal general aplicada a una configuración da lugar a una deformación de la misma, ya que en general no mantiene constantes los cocientes de las distancias entre los puntos de la configuración, lo que implica un cambio en su forma (véase ejemplo 22). No obstante, los resultados que se van a obtener en esta sección, servirán de base para presentar los casos donde se suponen transformaciones de similitud (rotaciones, traslaciones y escalamientos) y transformaciones rígidas, que son ampliamente extendidas y utilizadas tanto en el análisis de formas como en análisis Procrustes.

El modelo (3.1) tomando $\phi_i(x_{i+1j}) = A_i x_{i+1j}$ quedaría de la forma:

$$\begin{aligned}
 x_{1j} &= \mu_{\xi_{1j}} + \varepsilon_{1j} & j &= 1, \dots, n_1 \\
 A_1 x_{2j} &= \mu_{\xi_{2j}} + \varepsilon_{2j} & j &= 1, \dots, n_2 \\
 &\vdots & &\vdots \\
 A_{r-1} x_{rj} &= \mu_{\xi_{rj}} + \varepsilon_{rj} & j &= 1, \dots, n_r \\
 &\vdots & &\vdots \\
 A_{M-1} x_{Mj} &= \mu_{\xi_{Mj}} + \varepsilon_{Mj} & j &= 1, \dots, n_M
 \end{aligned} \tag{3.6}$$

donde para todo $i = 1, \dots, M$ y $j = 1, \dots, n_i$, $A_{i-1} \in \mathbb{R}^{d \times d}$ es una matriz no singular, con $A_0 = \mathbf{I}_d$, y $\{\varepsilon_{ij}\}$ son independientes con función de densidad f_i

A partir de (3.6), se observa que la función de densidad de x_{ij} , condicionada por A_{i-1} , ξ_{ij} , y $\{\mu_i\}$ (para todo $i = 1, \dots, M$ y $j = 1, \dots, n_i$) es

$$f(x_{ij}) = f_i(A_{i-1}x_{ij} - \mu_{\xi_{ij}}) |A_{i-1}|, \tag{3.7}$$

donde $|A|$ denota el valor absoluto del determinante de A .

3.5.1. Construcción de la verosimilitud de los datos

Para construir la verosimilitud de los datos, conocidas las matrices de afinidad y de emparejamientos, deberemos diferenciar y agrupar los puntos de las configuraciones según estén o no emparejados, y según el grado de emparejamiento.

A partir de (3.7), se construye la verosimilitud de todos los puntos de las configuraciones, diferenciando si los puntos están o no emparejados y, en el caso de estarlo, con qué orden de emparejamiento. Así, cada punto de una configuración puede no estar emparejado, formar parte de un emparejamiento doble, de uno triple,... De esta manera se calcula la aportación de cada subgrupo de puntos a la verosimilitud total.

Aportación a la verosimilitud de los puntos no emparejados

A partir de las propiedades del proceso de Poisson espacial, los $\{\mu_l\}$ que generan los puntos que no están emparejados, (aquellos que forman parte de la clase C_1), se distribuyen uniformemente sobre la región V (proposición 2).

De (3.7) se tiene que para todo $i = 1, \dots, M$ y para todo $j = 1, \dots, n_i$,

$$f(x_{ij}|\mu) = f_i(A_{i-1}x_{ij} - \mu) |A_{i-1}|.$$

Así,

$$\begin{aligned} f(x_{ij}) &= \int_V f(x_{ij}/\mu) f(\mu) d\mu = \\ &= |A_{i-1}| \int_V f_i(A_{i-1}x_{ij} - \mu) \frac{1}{v} d\mu \\ &= |A_{i-1}| \frac{1}{v} \int_V f_i(A_{i-1}x_{ij} - \mu) d\mu. \end{aligned}$$

Si denotamos como E_i^0 al conjunto de puntos de la configuración \mathbf{x}_i que no están emparejados, la aportación conjunta a la verosimilitud de todos los puntos no

emparejados será:

$$\begin{aligned} \prod_{i=1}^M \prod_{\{j, x_{ij} \in E_i^0\}} f(x_{ij}) &= \prod_{i=1}^M \prod_{\{j, x_{ij} \in E_i^0\}} \frac{1}{v} |A_{i-1}| \int_V f_i(A_{i-1}x_{ij} - \mu) d\mu \\ &= \left(\frac{1}{v}\right)^{n - \sum_{i=2}^M iL_i} \prod_{i=1}^M \prod_{\{j, x_{ij} \in E_i^0\}} |A_{i-1}| \int_V f_i(A_{i-1}x_{ij} - \mu) d\mu. \end{aligned} \quad (3.8)$$

Aportación a la verosimilitud de los puntos con emparejamiento doble

Los valores $\{\mu_l\}$ que generan los puntos que tienen un emparejamiento doble también se distribuyen uniformemente sobre la región V . Por otra parte, para todos los puntos $x_{i_1j_1}$ y $x_{i_2j_2}$ que están emparejados, es decir, para todo (i_1, i_2) tales que $M_{i_1, i_2}^{(2)} \neq 0$ y para todo (j_1, j_2) tales que $S_{j_1j_2}^{(i_1, i_2)} = 1$, la función de densidad conjunta de ambos, por la independencia entre los puntos, será:

$$\begin{aligned} f(x_{i_1j_1}, x_{i_2j_2}) &= \int_V f(x_{i_1j_1} | \mu) f(x_{i_2j_2} | \mu) f(\mu) d\mu = \\ &= \frac{1}{v} \int_V f_{i_1}(A_{i_1-1}x_{i_1j_1} - \mu) |A_{i_1-1}| f_{i_2}(A_{i_2-1}x_{i_2j_2} - \mu) |A_{i_2-1}| d\mu \\ &= \frac{1}{v} |A_{i_1-1}| |A_{i_2-1}| \int_V f_{i_1}(A_{i_1-1}x_{i_1j_1} - \mu) f_{i_2}(A_{i_2-1}x_{i_2j_2} - \mu) d\mu. \end{aligned}$$

Por lo tanto, la aportación conjunta a la verosimilitud de todos los puntos con emparejamientos dobles será:

$$\begin{aligned} &\prod_{\{(i_1, i_2) | M_{i_1, i_2}^{(2)} \neq 0\}} \prod_{\{(j_1, j_2) | S_{j_1j_2}^{(i_1, i_2)} = 1\}} f(x_{i_1j_1}, x_{i_2j_2}) \\ &= \left(\frac{1}{v}\right)^{L_2} \prod_{\substack{\{(i_1, i_2) | \\ M_{i_1, i_2}^{(2)} \neq 0\}}} \prod_{\substack{\{(j_1, j_2) | \\ S_{j_1j_2}^{(i_1, i_2)} = 1\}}} |A_{i_1-1}| |A_{i_2-1}| \cdot \\ &\quad \cdot \int_V f_{i_1}(A_{i_1-1}x_{i_1j_1} - \mu) f_{i_2}(A_{i_2-1}x_{i_2j_2} - \mu) d\mu. \end{aligned} \quad (3.9)$$

Aportación a la verosimilitud de los puntos con emparejamiento de orden k

En general, la aportación conjunta a la verosimilitud de todos los puntos con emparejamientos de orden k ($k = 2, \dots, M$), es:

$$\left(\frac{1}{v}\right)^{L_k} \prod_{\substack{\{(i_1, \dots, i_k) \mid \\ M_{i_1, \dots, i_k}^{(k)} \neq 0\}}} \prod_{\substack{\{(j_1, \dots, j_k) \mid \\ S_{j_1, \dots, j_k}^{(i_1, \dots, i_k)} = 1\}}} |A_{i_1-1}| \dots |A_{i_k-1}| \cdot \int_V f_{i_1}(A_{i_1-1}x_{i_1j_1} - \mu) \dots f_{i_k}(A_{i_k-1}x_{i_kj_k} - \mu) d\mu. \quad (3.10)$$

En el caso $k = M$, esta expresión se reduce a

$$\left(\frac{1}{v} |A_1| \dots |A_{M-1}|\right)^{L_M} \prod_{\substack{\{(j_1, \dots, j_k) \mid \\ S_{j_1, \dots, j_k}^{(1, \dots, M)} = 1\}}} \int_V f_1(x_{1j_1} - \mu) \dots f_M(A_{M-1}x_{Mj_M} - \mu) d\mu.$$

Expresión aproximada de la verosimilitud de los datos

Considerando la región $V \subset \mathbb{R}^d$ suficientemente grande con respecto al soporte de f_i , podemos aproximar V por \mathbb{R}^d , con lo que se consigue simplificar considerablemente las expresiones (3.8), (3.9) y (3.10).

Denotando por $e_i^{(k)}$ al número de emparejamientos de orden k que tiene la configuración i , se tiene que la aportación a la verosimilitud de los puntos no emparejados (3.8), se puede aproximar por,

$$\begin{aligned} \left(\frac{1}{v}\right)^{n - \sum_{i=2}^M iL_i} \prod_{i=1}^M \prod_{\{j, x_{ij} \in E_i^0\}} |A_{i-1}| \int_{\mathbf{R}^d} f_i(A_{i-1}x_{ij} - \mu) d\mu = \\ = \left(\frac{1}{v}\right)^{n - \sum_{i=2}^M iL_i} \prod_{i=1}^M |A_{i-1}|^{n_i - \sum_{l=2}^M e_i^{(l)}}, \end{aligned} \quad (3.11)$$

ya que $\int_{\mathbf{R}^d} f_i(A_{i-1}x_{ij} - \mu) d\mu = 1$. El exponente de $|A_{i-1}|$ representa el número de puntos de la configuración \mathbf{x}_i que no están emparejados.

De la misma manera, (3.9) puede aproximarse por,

$$\left(\frac{1}{v}\right)^{L_2} \prod_{\substack{\{(i_1, i_2)| \\ M_{i_1, i_2}^{(2)} \neq 0\}}} (|A_{i_1-1}| |A_{i_2-1}|)^{M_{i_1, i_2}^{(2)}} \cdot \prod_{\substack{\{(j_1, j_2)| \\ S_{j_1 j_2}^{(i_1, i_2)} = 1\}}} \int_{\mathbf{R}^d} f_{i_1}(A_{i_1-1}x_{i_1 j_1} - \mu) f_{i_2}(A_{i_2-1}x_{i_2 j_2} - \mu) d\mu. \quad (3.12)$$

Pero por (3.1) sabemos que $\varepsilon_{i_1 j_1}$ y $\varepsilon_{i_2 j_2}$ tienen como función de densidad f_{i_1} y f_{i_2} respectivamente (para todo j_1 y j_2), y que ambas distribuciones son independientes. Entonces, realizando un cambio de variable, la función de densidad de la variable $Z = \varepsilon_{i_1 j_1} - \varepsilon_{i_2 j_2}$ viene dada por la expresión,

$$g_{i_1, i_2}(z) = \int_{\mathbf{R}^d} f_{i_1}(w) f_{i_2}(w - z) dw, \quad (3.13)$$

por lo que la aproximación de la aportación de los emparejamientos de orden dos a la verosimilitud (3.12) puede expresarse como,

$$\left(\frac{1}{v}\right)^{L_2} \prod_{\substack{\{(i_1, i_2)| \\ M_{i_1, i_2}^{(2)} \neq 0\}}} (|A_{i_1-1}| |A_{i_2-1}|)^{M_{i_1, i_2}^{(2)}} \prod_{\substack{\{(j_1, j_2)| \\ S_{j_1 j_2}^{(i_1, i_2)} = 1\}}} g_{i_1, i_2}(A_{i_1-1}x_{i_1 j_1} - A_{i_2-1}x_{i_2 j_2}). \quad (3.14)$$

Para las expresiones de las aportaciones aproximadas del resto de los emparejamientos, utilizamos el siguiente resultado, que viene a ser la generalización de (3.13).

Proposición 4 *Dadas las variables aleatorias independientes $\varepsilon_{i_1 j_1}, \varepsilon_{i_2 j_2}, \dots, \varepsilon_{i_k j_k}$ ($k = 2, \dots, M$), con funciones de densidad $f_{i_1}, f_{i_2}, \dots, f_{i_k}$ se tiene que la distribución conjunta de $(Z_2, \dots, Z_k) = (\varepsilon_{i_1 j_1} - \varepsilon_{i_2 j_2}, \varepsilon_{i_1 j_1} - \varepsilon_{i_3 j_3}, \dots, \varepsilon_{i_1 j_1} - \varepsilon_{i_k j_k})$ es,*

$$g_{i_1, i_2, i_3, \dots, i_k}(z_2, \dots, z_k) = \int_{\mathbf{R}^d} f_{i_1}(w) f_{i_2}(w - z_2) f_{i_3}(w - z_3) \dots f_{i_k}(w - z_k) dw. \quad (3.15)$$

Demostración. Dado $k = 2, \dots, M$, hacemos el cambio de variable

$$\begin{array}{ll} w = \varepsilon_{i_1 j_1} & \varepsilon_{i_1 j_1} = w \\ z_2 = \varepsilon_{i_1 j_1} - \varepsilon_{i_2 j_2} & \Rightarrow \varepsilon_{i_2 j_2} = w - z_2 \\ \dots & \dots \\ z_k = \varepsilon_{i_1 j_1} - \varepsilon_{i_k j_k} & \varepsilon_{i_k j_k} = w - z_k \end{array}$$

por lo que el jacobiano $|J| = 1$. Así $g(w, z_2, \dots, z_k) = f(w, w - z_2, \dots, w - z_k) = f_{i_1}(w) \cdot f_{i_2}(w - z_2) \cdot \dots \cdot f_{i_k}(w - z_k)$. Por lo tanto

$$g_{i_1, i_2, i_3, \dots, i_k}(z_2, \dots, z_k) = \int_{\mathbf{R}^d} f_{i_1}(w) f_{i_2}(w - z_2) f_{i_3}(w - z_3) \dots f_{i_k}(w - z_k) dw,$$

como se quería demostrar. ■

Entonces, a partir de (3.10) y de (3.15), la aproximación de la aportación de la verosimilitud aportada por los puntos con emparejamientos de orden $k = 2, \dots, M$, tendrá la expresión,

$$\begin{aligned} & \left(\frac{1}{v}\right)^{L_k} \prod_{\substack{\{(i_1, \dots, i_k) \mid \\ M_{i_1 \dots i_k}^{(k)} \neq 0\}}} (|A_{i_1-1}| \dots |A_{i_k-1}|)^{M_{i_1 \dots i_k}^{(k)}} \cdot \\ & \prod_{\substack{\{(j_1 \dots j_k) \mid \\ S_{j_1 \dots j_k}^{(i_1 \dots i_k)} = 1\}}} g_{i_1, i_2, \dots, i_k}(A_{i_1-1}x_{i_1 j_1} - A_{i_2-1}x_{i_2 j_2}, \dots, A_{i_1-1}x_{i_1 j_1} - A_{i_k-1}x_{i_k j_k}). \end{aligned} \quad (3.16)$$

Se puede observar que (3.14) es el caso particular de (3.16) para $k = 2$. Por lo tanto, la aproximación de la verosimilitud de todos los puntos, se obtendrá a través del producto de (3.11) y de todas las expresiones (3.16) para los valores $k = 2, \dots, M$.

3.5.2. Emparejamiento de $M = 4$ configuraciones con errores normales

Como un caso particular de transformaciones lineales generales vamos a concretar estos resultados en el caso en el que los $\{\varepsilon_{ij}\}$ tienen distribución normal, tomando $M = 4$ configuraciones con el único objetivo de simplificar las expresiones.

El modelo (3.6) en el caso de cuatro configuraciones quedaría reducido a,

$$\begin{aligned} x_{1j} &= \mu_{\xi_{1j}} + \varepsilon_{1j} & j &= 1, \dots, n_1 \\ A_1 x_{2j} &= \mu_{\xi_{2j}} + \varepsilon_{2j} & j &= 1, \dots, n_2 \\ A_2 x_{3j} &= \mu_{\xi_{3j}} + \varepsilon_{3j} & j &= 1, \dots, n_3 \\ A_3 x_{4j} &= \mu_{\xi_{4j}} + \varepsilon_{4j} & j &= 1, \dots, n_4 \end{aligned}$$

donde para todo $i = 1, \dots, 4$ y $j = 1, \dots, n_i$, A_{i-1} son matrices no singulares y los $\{\varepsilon_{ij}\}$ son independientes con distribución $\mathbf{N}_d(0, \sigma^2 \mathbf{I}_d)$, con $d \geq 2$.

En este caso, los parámetros del modelo son:

- a) Las matrices de afinidad de órdenes dos, tres y cuatro: $\mathbf{M}^{(2)}$, $\mathbf{M}^{(3)}$ y $\mathbf{M}^{(4)}$.
- b) Las matrices de emparejamientos de órdenes dos, tres y cuatro: $\mathbf{S}^{(1,2)}$, $\mathbf{S}^{(1,3)}$, $\mathbf{S}^{(1,4)}$, $\mathbf{S}^{(2,3)}$, $\mathbf{S}^{(2,4)}$, $\mathbf{S}^{(3,4)}$, $\mathbf{S}^{(1,2,3)}$, $\mathbf{S}^{(1,2,4)}$, $\mathbf{S}^{(2,3,4)}$ y $\mathbf{S}^{(1,2,3,4)}$.
- c) Las matrices no singulares que definen las transformaciones lineales: A_1 , A_2 y A_3 .
- d) La varianza de la distribución normal de los errores: σ^2 .

La distribución conjunta de las matrices de afinidad particularizada en el caso de cuatro configuraciones, se puede deducir de (3.4), obteniendo la expresión,

$$p(\mathbf{M}^{(2)}, \mathbf{M}^{(3)}, \mathbf{M}^{(4)}) \propto \frac{\left(\frac{\rho_2}{\lambda v}\right)^{L_2} \left(\frac{\rho_3}{\lambda v}\right)^{L_3} \left(\frac{\rho_4}{\lambda v}\right)^{L_4}}{(\lambda v)^{L_3+2L_4} (n - 2L_2 - 3L_3 - 4L_4)! (5 + L_2)! (3 + L_3)! L_4!}. \quad (3.17)$$

A partir de (3.5) se construye la expresión de la distribución conjunta de las matrices de emparejamientos de orden dos, dada la matriz de afinidad de orden 2,

$$p(\mathbf{S}^{(1,2)}, \mathbf{S}^{(1,3)}, \mathbf{S}^{(1,4)}, \mathbf{S}^{(2,3)}, \mathbf{S}^{(2,4)}, \mathbf{S}^{(3,4)} \mid \mathbf{M}^{(2)}) = \frac{1}{\prod_{\{(i_1, i_2) \mid \mathbf{M}_{i_1, i_2}^{(2)} \neq 0\}} \binom{n_{i_1}}{\mathbf{M}_{i_1, i_2}^{(2)}} \binom{n_{i_2}}{\mathbf{M}_{i_1, i_2}^{(3)}} \left(\mathbf{M}_{i_1, i_2}^{(2)}!\right)}, \quad (3.18)$$

la distribución conjunta de las matrices de emparejamientos de orden tres, dada la matriz de afinidad de orden 3,

$$p(\mathbf{S}^{(1,2,3)}, \mathbf{S}^{(1,2,4)}, \mathbf{S}^{(2,3,4)} \mid \mathbf{M}^{(3)}) = \frac{1}{\prod_{\{(i_1, i_2, i_3) \mid \mathbf{M}_{i_1, i_2, i_3}^{(3)} \neq 0\}} \binom{n_{i_1}}{\mathbf{M}_{i_1, i_2, i_3}^{(3)}} \binom{n_{i_2}}{\mathbf{M}_{i_1, i_2, i_3}^{(3)}} \binom{n_{i_3}}{\mathbf{M}_{i_1, i_2, i_3}^{(3)}} \left(\mathbf{M}_{i_1, i_2, i_3}^{(3)}!\right)^2}, \quad (3.19)$$

y la distribución de la matriz de emparejamientos de orden cuatro, dada la matriz de afinidad de orden 4, o equivalentemente, dado L_4 , pues esta es la información recogida en la matriz $\mathbf{M}^{(4)}$,

$$p(\mathbf{S}^{(1,2,3,4)} \mid \mathbf{M}^{(4)}) = \frac{1}{\binom{n_1}{L_4} \binom{n_2}{L_4} \binom{n_3}{L_4} \binom{n_4}{L_4} (L_4!)^3}. \quad (3.20)$$

Por lo tanto la distribución conjunta de todas las matrices de afinidad y de emparejamientos, se construirá mediante el producto de (3.17), (3.18), (3.19) y (3.20).

Verosimilitud de los datos en caso de normalidad

Como caso particular de la aportación aproximada a la verosimilitud de los puntos no emparejados (3.11), para $M = 4$ se obtiene la expresión,

$$\left(\frac{1}{v}\right)^{n-2L_2-3L_3-4L_4} |A_1|^{n_2-\sum_{k=2}^M e_2^{(k)}} |A_2|^{n_3-\sum_{k=2}^M e_3^{(k)}} |A_3|^{n_4-\sum_{k=2}^M e_4^{(k)}}, \quad (3.21)$$

donde

$$\begin{aligned} \sum_{k=2}^M e_2^{(k)} &= \mathbf{M}_{12}^{(2)} + \mathbf{M}_{23}^{(2)} + \mathbf{M}_{24}^{(2)} + \mathbf{M}_{123}^{(3)} + \mathbf{M}_{124}^{(3)} + \mathbf{M}_{234}^{(3)} + \mathbf{M}_{1234}^{(4)}, \\ \sum_{k=2}^M e_3^{(k)} &= \mathbf{M}_{13}^{(2)} + \mathbf{M}_{23}^{(2)} + \mathbf{M}_{34}^{(2)} + \mathbf{M}_{123}^{(3)} + \mathbf{M}_{134}^{(3)} + \mathbf{M}_{234}^{(3)} + \mathbf{M}_{1234}^{(4)}, \\ \sum_{k=2}^M e_4^{(k)} &= \mathbf{M}_{14}^{(2)} + \mathbf{M}_{24}^{(2)} + \mathbf{M}_{34}^{(2)} + \mathbf{M}_{124}^{(3)} + \mathbf{M}_{134}^{(3)} + \mathbf{M}_{234}^{(3)} + \mathbf{M}_{1234}^{(4)}, \end{aligned}$$

representan el número de puntos emparejados en las configuraciones \mathbf{x}_2 , \mathbf{x}_3 y \mathbf{x}_4 , respectivamente.

Para encontrar las expresiones de las aportaciones de los puntos con emparejamientos dobles, emparejamientos triples y emparejamientos de orden cuatro, bajo hipótesis de normalidad y deducirlas de las expresiones (3.16), nos basamos en el siguiente resultado.

Proposición 5 Sean Y_1, Y_2, Y_3 y Y_4 vectores aleatorios independientes con distribución $\mathbf{N}_d(0, \sigma^2 \mathbf{I}_d)$. Se tiene que:

- (i) Para todo $i, j \in \{1, 2, 3, 4\}$, las diferencias $Z_{ij} = Y_i - Y_j$, tienen como función de densidad

$$g_{i,j}(z) = \left(\frac{1}{\sigma\sqrt{2}} \right)^d \varphi_d\left(\frac{z}{\sigma\sqrt{2}}\right),$$

donde $\varphi_d(z)$ es la función de densidad de la distribución $\mathbf{N}_d(0, \mathbf{I}_d)$.

- (ii) Para todo $j_1, j_2 \in \{2, 3, 4\}$ la distribución conjunta de $W_{j_1 j_2} = (Z_{1j_1}, Z_{1j_2}) = (Y_1 - Y_{j_1}, Y_1 - Y_{j_2})$ tiene como función de densidad

$$g_{1,j_1,j_2}(z_{j_1}, z_{j_2}) = \left(\frac{1}{\sigma^2\sqrt{3}} \right)^d \varphi_{2d}\left(\frac{z_{j_1}}{\sigma\sqrt{2}}, \frac{\sqrt{6}}{3\sigma}(z_{j_2} - \frac{1}{2}z_{j_1})\right), \quad (3.22)$$

donde $\varphi_{2d}(z, v)$ es la función de densidad de la distribución $\mathbf{N}_{2d}(0, \mathbf{I}_{2d})$.

- (iii) La distribución conjunta de $W = (Z_{12}, Z_{13}, Z_{14}) = (Y_1 - Y_2, Y_1 - Y_3, Y_1 - Y_4)$ tiene función de densidad

$$g_{1,2,3,4}(z_2, z_3, z_4) = \left(\frac{1}{2\sigma^3} \right)^d \varphi_{3d}\left(\frac{z_2}{\sigma\sqrt{2}}, \frac{1}{\sigma\sqrt{6}}(-z_2 + 2z_3), \frac{1}{\sigma\sqrt{12}}(-z_2 - z_3 + 3z_4)\right),$$

donde $\varphi_{3d}(z, v, w)$ es la función de densidad de la distribución $\mathbf{N}_{3d}(0, \mathbf{I}_{3d})$.

Demostración.

- (i) Al ser independientes Y_i y Y_j para todo $i, j \in \{1, 2, 3, 4\}$, el vector aleatorio $Z_{ij} = Y_i - Y_j \sim \mathbf{N}_d(0, 2\sigma^2\mathbf{I}_d)$. Entonces Z_{ij} se puede expresar como $Z_{ij} = \sigma\sqrt{2}Y$, donde $Y \sim \mathbf{N}_d(0, \mathbf{I}_d)$ con función de densidad $\varphi_d(y)$. Así, haciendo el cambio de variable, $Y = Z_{ij}/\sigma\sqrt{2}$, se tiene que el jacobiano $|J| = \left(\frac{1}{\sigma\sqrt{2}}\right)^d$ por lo que para todo $i, j \in \{1, 2, 3, 4\}$, la función de densidad de Z_{ij} , será

$$g(z) = \left(\frac{1}{\sigma\sqrt{2}} \right)^d \varphi_d\left(\frac{z}{\sigma\sqrt{2}}\right),$$

quedando demostrado (i).

- (ii) De la proposición 5(i) se tiene que Z_{1j_1} y Z_{1j_2} se distribuyen $\mathbf{N}_d(0, 2\sigma^2\mathbf{I}_d)$. Pero no son independientes, ya que la matriz de varianzas-covarianzas de $W_{j_1 j_2} = (Z_{1j_1}, Z_{1j_2})$ es:

$$\mathbf{V}(W_{j_1 j_2}) = \Sigma = \sigma^2 \begin{pmatrix} 2\mathbf{I}_d & \mathbf{I}_d \\ \mathbf{I}_d & 2\mathbf{I}_d \end{pmatrix}.$$

Entonces, la distribución de $W_{j_1 j_2} = (Z_{1j_1}, Z_{1j_2})$ es $\mathbf{N}_{2d}(0, \Sigma)$. Por otro lado, para escribir la función de densidad de $W_{j_1 j_2}$ en términos de la función de densidad de la distribución $\mathbf{N}_{2d}(0, \mathbf{I}_{2d})$, utilizamos la descomposición de Cholesky de la matriz Σ , es decir, utilizamos que existe una matriz A tal que $\Sigma = A^T A$, y por lo tanto, $W_{j_1 j_2} = A^T Y_{2d}$ donde $Y_{2d} \sim \mathbf{N}_{2d}(0, \mathbf{I}_{2d})$. Esta matriz es,

$$A = \sigma \begin{pmatrix} \sqrt{2}\mathbf{I}_d & \frac{\sqrt{2}}{2}\mathbf{I}_d \\ \mathbf{0}_d & \frac{\sqrt{6}}{2}\mathbf{I}_d \end{pmatrix} = \sigma\sqrt{2} \begin{pmatrix} \mathbf{I}_d & \frac{1}{2}\mathbf{I}_d \\ \mathbf{0}_d & \frac{\sqrt{3}}{2}\mathbf{I}_d \end{pmatrix}.$$

Así, calculamos la densidad de $W_{j_1 j_2}$ haciendo el cambio de variable $Y_{2d} = (A^T)^{-1}W_{j_1 j_2}$ y entonces,

$$g(w) = g_{1,j_1,j_2}(z_{j_1}, z_{j_2}) = \varphi_{2d}((A^T)^{-1}w) \left| (A^T)^{-1} \right| = \varphi_{2d}((A^T)^{-1}(z_{j_1}, z_{j_2})) \left| (A^T)^{-1} \right|.$$

Pero $(A^T)^{-1} = \frac{1}{\sigma} \begin{pmatrix} \frac{\sqrt{2}}{2}\mathbf{I}_d & 0 \\ -\frac{\sqrt{6}}{6}\mathbf{I}_d & \frac{\sqrt{6}}{3}\mathbf{I}_d \end{pmatrix}$, y su determinante $\left| (A^T)^{-1} \right| = \left(\frac{1}{\sigma^2\sqrt{3}} \right)^d$.

Por lo tanto,

$$g_{1,j_1,j_2}(z_{j_1}, z_{j_2}) = \varphi_{2d} \left(\frac{\sqrt{2}}{2\sigma} z_{j_1}, -\frac{\sqrt{6}}{6\sigma} z_{j_1} + \frac{\sqrt{6}}{3\sigma} z_{j_2} \right) \left(\frac{1}{\sigma^2\sqrt{3}} \right)^d,$$

que coincide con la expresión (3.22).

- (iii) Por último, sabemos que Z_{12} , Z_{13} y Z_{14} se distribuyen $\mathbf{N}_d(0, 2\sigma^2\mathbf{I}_d)$. Entonces, la matriz de varianzas covarianzas de $W = (Z_{12}, Z_{13}, Z_{14})$ es

$$\mathbf{V}(W) = \Sigma_W = \sigma^2 \begin{pmatrix} 2\mathbf{I}_d & \mathbf{I}_d & \mathbf{I}_d \\ \mathbf{I}_d & 2\mathbf{I}_d & \mathbf{I}_d \\ \mathbf{I}_d & \mathbf{I}_d & 2\mathbf{I}_d \end{pmatrix},$$

por lo que la distribución de $W = (Z_{12}, Z_{13}, Z_{14})$ es $\mathbf{N}_{3d}(0, \Sigma_W)$.

De la misma manera que en la demostración de (ii), utilizamos la descomposición de Cholesky de la matriz Σ_W , con el objetivo de expresar la densidad de W en términos de la función de densidad de la distribución normal estándar 3d-dimensional. Es inmediato que la descomposición de Cholesky de $\Sigma_W = T^T T$ con

$$T = \sigma \begin{pmatrix} \sqrt{2}\mathbf{I}_d & \frac{\sqrt{2}}{2}\mathbf{I}_d & \frac{\sqrt{2}}{2}\mathbf{I}_d \\ \mathbf{0}_d & \sqrt{\frac{3}{2}}\mathbf{I}_d & \frac{1}{2}\sqrt{\frac{2}{3}}\mathbf{I}_d \\ \mathbf{0}_d & \mathbf{0}_d & \sqrt{\frac{4}{3}}\mathbf{I}_d \end{pmatrix}.$$

Haciendo el cambio de variable $W = T^T Y_{3d}$ donde $Y_{3d} \sim \mathbf{N}_{3d}(0, \mathbf{I}_{3d})$ y utilizando que

$$(T^T)^{-1} = \frac{1}{\sigma} \begin{pmatrix} \frac{\sqrt{2}}{2} \mathbf{I}_d & \mathbf{0}_d & \mathbf{0}_d \\ -\frac{1}{\sqrt{6}} \mathbf{I}_d & \sqrt{\frac{2}{3}} \mathbf{I}_d & \mathbf{0}_d \\ -\frac{1}{\sqrt{12}} \mathbf{I}_d & -\frac{1}{\sqrt{12}} \mathbf{I}_d & \sqrt{\frac{3}{4}} \mathbf{I}_d \end{pmatrix},$$

con determinante $\left| (T^T)^{-1} \right| = \left(\frac{1}{2\sigma^2} \right)^d$, se obtiene que la función de densidad de $W = (Z_{12}, Z_{13}, Z_{14})$ es

$$g(w) = g_{1,2,3,4}(z_2, z_3, z_4) = \left(\frac{1}{2\sigma^3} \right)^d \varphi_{3d} \left(\frac{z_2}{\sigma\sqrt{2}}, \frac{1}{\sigma\sqrt{6}}(-z_2 + 2z_3), \frac{1}{\sigma\sqrt{12}}(-z_2 - z_3 + 3z_4) \right).$$

■

La proposición 5(i) nos lleva a deducir, a partir de (3.16), que la aportación aproximada a la verosimilitud de los emparejamientos dobles en el caso de cuatro configuraciones y bajo hipótesis de normalidad tiene la expresión,

$$\left(\frac{1}{v} \right)^{L_2} \prod_{\substack{\{(i_1, i_2) | \\ M_{i_1, i_2}^{(2)} \neq 0\}}} (|A_{i_1-1}| |A_{i_2-1}|)^{M_{i_1, i_2}^{(2)}} \prod_{\substack{\{(j_1, j_2) | \\ S_{j_1 j_2}^{(i_1, i_2)} = 1\}}} \left(\frac{1}{\sigma\sqrt{2}} \right)^d \varphi_d \left(\frac{A_{i_1-1}x_{i_1 j_1} - A_{i_2-1}x_{i_2 j_2}}{\sigma\sqrt{2}} \right). \quad (3.23)$$

Por otro lado, a partir de (3.16) en el caso $k = 3$, se deduce que la aportación aproximada a la verosimilitud de los emparejamientos triples, en el caso de $M = 4$ tiene la expresión,

$$\left(\frac{1}{v} \right)^{L_3} \prod_{\substack{\{(i_1, i_2, i_3) | \\ M_{i_1, i_2, i_3}^{(3)} \neq 0\}}} |A_{i_1-1} A_{i_2-1} A_{i_3-1}|^{M_{i_1, i_2, i_3}^{(3)}} \cdot \prod_{\substack{\{(j_1, j_2, j_3) | \\ S_{j_1, j_2, j_3}^{(i_1, i_2, i_3)} = 1\}}} g_{i_1, i_2, i_3}(A_{i_1-1}x_{i_1 j_1} - A_{i_2-1}x_{i_2 j_2}, A_{i_1-1}x_{i_1 j_1} - A_{i_3-1}x_{i_3 j_3}).$$

Además en el caso de normalidad, de la proposición 5(ii), esta aportación tiene

la forma

$$\begin{aligned} & \left(\frac{1}{v}\right)^{L_3} \prod_{\substack{\{(i_1, i_2, i_3)\} \\ M_{i_1, i_2, i_3}^{(3)} \neq 0}} |A_{i_1-1} A_{i_2-1} A_{i_3-1}|^{M_{i_1, i_2, i_3}^{(3)}} \prod_{\substack{\{(j_1, j_2, j_3)\} \\ S_{j_1, j_2, j_3}^{(1, 2, 3)} = 1}} \left(\frac{1}{\sigma^2 \sqrt{3}}\right)^d \cdot \\ & \cdot \varphi_{2d} \left(\frac{A_{i_1-1} x_{i_1 j_1} - A_{i_2-1} x_{i_2 j_2}}{\sigma \sqrt{2}}, \right. \\ & \left. \frac{\sqrt{6}}{6\sigma} (A_{i_1-1} x_{i_1 j_1} + A_{i_2-1} x_{i_2 j_2} - 2A_{i_3-1} x_{i_3 j_3}) \right). \end{aligned} \quad (3.24)$$

Por último, la aportación aproximada a la verosimilitud de los emparejamientos de orden 4 a partir de (3.16) y de la proposición 5(iii), es de la forma

$$\begin{aligned} & \left(\frac{1}{v} |A_1 A_2 A_3|\right)^{L_4} \prod_{\substack{\{(j_1, j_2, j_3, j_4)\} \\ S_{j_1, j_2, j_3, j_4}^{(1, 2, 3, 4)} = 1}} \left(\frac{1}{2\sigma^3}\right)^d \varphi_{3d} \left(\frac{x_{1j_1} - A_1 x_{2j_2}}{\sigma \sqrt{2}}, \right. \\ & \left. \frac{x_{1j_1} + A_1 x_{2j_2} - 2A_2 x_{3j_3}}{\sigma \sqrt{6}}, \frac{x_{1j_1} + A_1 x_{2j_2} + A_2 x_{3j_3} - 3A_3 x_{4j_4}}{\sigma \sqrt{12}} \right). \end{aligned} \quad (3.25)$$

Por lo tanto, bajo hipótesis de normalidad, la aproximación de la verosimilitud de los puntos de cuatro configuraciones será el producto de las expresiones (3.21), (3.23), (3.24) y (3.25).

3.5.3. Inferencia bayesiana en el caso de emparejamientos conocidos de $M = 4$ configuraciones

Partiendo de $M = 4$ configuraciones en \mathbb{R}^d , vamos a añadir la restricción de que todas ellas tienen el mismo número de puntos, es decir, $n_i = m$ para todo $i = 1, 2, 3, 4$. Además, vamos a suponer que se sabe que hay m emparejamientos cuádruples y también cuáles son los puntos que están emparejados. Así, los parámetros del modelo se reducen a aquellos con los que se definen las transformaciones entre configuraciones, es decir, las matrices A_1, A_2, A_3 y el parámetro de precisión σ^2 y sobre ellos aplicaremos la metodología bayesiana.

Sin pérdida de generalidad, supongamos que $\forall j = 1, \dots, m$ los puntos x_{1j}, x_{2j}, x_{3j} y x_{4j} forman los m emparejamientos de orden cuatro, es decir, $\xi_{1j} = \xi_{2j} = \xi_{3j} = \xi_{4j}$. En esta situación, las matrices de afinidad y de emparejamientos son conocidas. En particular, las matrices de afinidad de orden 2 y 3 son cero (por tanto también las de emparejamientos de este orden), y la matriz de afinidad de orden 4 toma el valor m en la coordenada $(1,2,3,4)$ y permutaciones de ella, y cero en el resto, es decir,

$$M_{ijkl}^{(4)} = \begin{cases} m & \text{si } \{i, j, k, l\} \in \{\pi(1, 2, 3, 4)\} \\ 0 & \text{resto} \end{cases}$$

donde $\pi(1, 2, 3, 4)$ denota las permutaciones de los números $\{1, 2, 3, 4\}$.

La matriz de emparejamientos de orden cuatro $\mathbf{S}^{(1,2,3,4)}$, tendrá dimensiones $m \times m \times m \times m$, y tomará el valor uno en las coordenadas $\mathbf{S}_{j,j,j,j}^{(1,2,3,4)}$ para todo $j = 1, \dots, m$ y cero en el resto, es decir,

$$S_{ijkl}^{(1,2,3,4)} = \begin{cases} 1 & \text{si } i = j = k = l \in \{1, \dots, m\} \\ 0 & \text{resto} \end{cases}$$

En primer lugar, calculamos la expresión de la verosimilitud de los datos y a continuación calcularemos las distribuciones a posteriori de los parámetros asumiendo ciertas distribuciones a priori.

La verosimilitud de los datos se reducirá a la expresión asociada a los emparejamientos de orden cuatro (3.25) que, particularizada en esta situación quedaría de la forma,

$$\begin{aligned} p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \mid A_1, A_2, A_3, \sigma^2) &\propto \left(\frac{1}{v} |A_1 A_2 A_3| \right)^m \prod_{j=1}^m \left(\frac{1}{2\sigma^3} \right)^d \cdot \\ &\varphi_{3d} \left(\frac{x_{1j} - A_1 x_{2j}}{\sigma\sqrt{2}}, \frac{x_{1j} + A_1 x_{2j} - 2A_2 x_{3j}}{\sigma\sqrt{6}}, \frac{x_{1j} + A_1 x_{2j} + A_2 x_{3j} - 3A_3 x_{4j}}{\sigma\sqrt{12}} \right) \\ &\propto \left(\frac{1}{2\sigma^3} \right)^{md} \exp \left\{ -\frac{1}{2} \left(\sum_{j=1}^m \left\| \frac{x_{1j} - A_1 x_{2j}}{\sigma\sqrt{2}} \right\|^2 \right. \right. \\ &\quad \left. \left. + \sum_{j=1}^m \left\| \frac{x_{1j} + A_1 x_{2j} - 2A_2 x_{3j}}{\sigma\sqrt{6}} \right\|^2 + \sum_{j=1}^m \left\| \frac{x_{1j} + A_1 x_{2j} + A_2 x_{3j} - 3A_3 x_{4j}}{\sigma\sqrt{12}} \right\|^2 \right) \right\}. \end{aligned}$$

Con el objetivo de conseguir una expresión más manejable, denotamos por

$$\begin{aligned} B_{1j} &= \frac{1}{\sqrt{2}}(x_{1j} - A_1 x_{2j}), \\ B_{2j} &= \frac{1}{\sqrt{6}}(x_{1j} + A_1 x_{2j} - 2A_2 x_{3j}), \\ B_{3j} &= \frac{1}{\sqrt{12}}(x_{1j} + A_1 x_{2j} + A_2 x_{3j} - 3A_3 x_{4j}). \end{aligned}$$

Así la verosimilitud se puede expresar como:

$$\begin{aligned} p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \mid A_1, A_2, A_3, \sigma^2) &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{3}{2}md} |A_1 A_2 A_3|^m \\ &\exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{j=1}^m \|B_{1j}\|^2 + \sum_{j=1}^m \|B_{2j}\|^2 + \sum_{j=1}^m \|B_{3j}\|^2 \right) \right\} \end{aligned} \quad (3.26)$$

Distribución a posteriori de los parámetros Se sume que la distribución a priori de la varianza σ^2 es una gamma inversa $\sigma^{-2} \sim G(\alpha, \beta)$, $\alpha, \beta > 0$, cuya función de densidad es,

$$p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-(\alpha+1)} \exp\{-\beta/\sigma^2\}.$$

A partir de (3.26) se obtiene que la distribución a posteriori de σ^2 es de la forma,

$$\begin{aligned} p(\sigma^2 \mid \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, A_1, A_2, A_3) &\propto \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-(\alpha+1)} \exp\{-\beta/\sigma^2\} \cdot \left(\frac{1}{\sigma^2}\right)^{\frac{3}{2}md} \\ &\exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^m (\|B_{1j}\|^2 + \|B_{2j}\|^2 + \|B_{3j}\|^2) \right\} = \\ &= (\sigma^2)^{-(\alpha+\frac{3}{2}md+1)} \exp \left\{ -\frac{1}{\sigma^2} \left(\beta + \frac{1}{2} \sum_{j=1}^m (\|B_{1j}\|^2 + \|B_{2j}\|^2 + \|B_{3j}\|^2) \right) \right\}. \end{aligned}$$

Por lo tanto, la distribución a posteriori de σ^2 es una gamma inversa $\sigma^{-2} \sim$

$G(\alpha^*, \beta^*)$ donde

$$\begin{aligned}\alpha^* &= \alpha + \frac{3}{2}md, \\ \beta^* &= \beta + \left\{ \frac{1}{2} \sum_{j=1}^m \|x_{1j} - A_1 x_{2j}\|^2 + \frac{1}{6} \sum_{j=1}^m \|x_{1j} + A_1 x_{2j} - 2A_2 x_{3j}\|^2 \right. \\ &\quad \left. + \frac{1}{12} \sum_{j=1}^m \|x_{1j} + A_1 x_{2j} + A_2 x_{3j} - 3A_3 x_{4j}\|^2 \right\}.\end{aligned}\quad (3.27)$$

A continuación estudiaremos las distribuciones a posteriori de las matrices de las transformaciones lineales, suponiendo distribución a priori Matriz Normal.

Definición 32 Una matriz $X \in \mathbb{R}^{n \times p}$ se dice que tiene distribución Matriz Normal $X \sim N(M, \Phi \otimes \Sigma)$, donde \otimes es el producto de Kronecker, $M \in \mathbb{R}^{n \times p}$ y $\Sigma_{p \times p}, \Phi_{n \times n} > 0$, si su función de densidad es

$$p(X \mid M, \Phi, \Sigma) = (2\pi)^{-\frac{np}{2}} |\Phi|^{-p/2} |\Sigma|^{-\frac{n}{2}} \exp \left(-\frac{1}{2} \text{traza} \Phi^{-1} (X - M) \Sigma^{-1} (X - M)^T \right).$$

El vector aleatorio $\text{vec}(X^T) = (x_1^T, x_2^T, \dots, x_n^T)^T$ donde x_i^T representa la fila i -ésima de la matriz traspuesta de X , tiene distribución normal $n \times p$ -dimensional de vector de media $\text{vec}(M)$ y matriz de varianzas-covarianzas $\Omega = \Phi \otimes \Sigma$.

Se asume que las distribuciones a priori de las matrices A_1 , A_2 y A_3 son Matriz Normales $N(M, \Phi \otimes \Sigma)$, con $M = \mathbf{0}_{n \times p}$, $\Phi = I_n$ y $\Sigma = I_p$, es decir,

$$p(A_i) \propto \exp \left\{ -\frac{1}{2} \text{traza}(A_i A_i^T) \right\}, \quad i = 1, 2, 3.$$

Entonces, a partir (3.26) la distribución a posteriori de A_1 es

$$\begin{aligned}p(A_1 \mid \dots) &\propto |A_1|^m \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{j=1}^m \|B_{1j}\|^2 + \sum_{j=1}^m \|B_{2j}\|^2 + \sum_{j=1}^m \|B_{3j}\|^2 \right) \right. \\ &\quad \left. - \frac{1}{2} \text{traza}(A_1 A_1^T) \right\}.\end{aligned}\quad (3.28)$$

Se puede simplificar la expresión en términos de la traza de una matriz.

$$\begin{aligned}\sum_{j=1}^m \|B_{1j}\|^2 &= \sum_{j=1}^m \left\| \frac{x_{1j} - A_1 x_{2j}}{\sqrt{2}} \right\|^2 = \frac{1}{2} \sum_{j=1}^m (x_{1j} - A_1 x_{2j})^T (x_{1j} - A_1 x_{2j}) \\ &= \frac{1}{2} \sum_{j=1}^m (\|x_{1j}\|^2 + \|A_1 x_{2j}\|^2 - 2x_{1j}^T A_1 x_{2j}).\end{aligned}\quad (3.29)$$

Análogamente,

$$\begin{aligned}\sum_{j=1}^m \|B_{2j}\|^2 &= \sum_{j=1}^m \left\| \frac{x_{1j} + A_1 x_{2j} - 2A_2 x_{3j}}{\sqrt{6}} \right\|^2 \\ &= \frac{1}{6} \sum_{j=1}^m (\|A_1 x_{2j}\|^2 + \|2A_2 x_{3j} - x_{1j}\|^2 - 2(2A_2 x_{3j} - x_{1j})^T A_1 x_{2j}).\end{aligned}\quad (3.30)$$

Por último,

$$\begin{aligned}\sum_{j=1}^m \|B_{3j}\|^2 &= \sum_{j=1}^m \left\| \frac{x_{1j} + A_1 x_{2j} + A_2 x_{3j} - 3A_3 x_{4j}}{\sqrt{12}} \right\|^2 \\ &= \frac{1}{12} \sum_{j=1}^m (\|A_1 x_{2j}\|^2 + \|3A_3 x_{4j} - A_2 x_{3j} - x_{1j}\|^2 \\ &\quad - 2(3A_3 x_{4j} - A_2 x_{3j} - x_{1j})^T A_1 x_{2j})\end{aligned}\quad (3.31)$$

Entonces, sustituyendo (3.29), (3.30) y (3.31) en la expresión (3.28), teniendo en cuenta que $a^T b = \text{traza}(ba^T)$ y simplificando, se llega a que

$$\begin{aligned}p(A_1 \mid \dots) &\propto |A_1|^m \exp \left\{ -\frac{1}{2} \left(\text{tr} \left(3 \sum_{j=1}^m \frac{A_1 x_{2j} x_{2j}^T A_1^T}{4\sigma^2} \right) \right. \right. \\ &\quad \left. \left. - \text{tr} \left(A_1 \frac{1}{2\sigma^2} \sum_{j=1}^m x_{2j} (x_{1j} + A_2 x_{3j} + A_3 x_{4j})^T \right) - \frac{1}{2} \text{tr} (A_1 A_1^T) \right) \right\}.\end{aligned}$$

Por lo tanto, la expresión de la distribución a posteriori de la matriz A_1 es

$$p(A_1 \mid \dots) \propto |A_1|^m \exp \left\{ -\frac{1}{2} \left[\text{tr} \left(A_1 \left(\frac{3}{4\sigma^2} \sum_{j=1}^m x_{2j} x_{2j}^T + I_d \right) A_1^T \right) - \text{tr} \left(\frac{1}{2\sigma^2} A_1 \sum_{j=1}^m x_{2j} (x_{1j} + A_2 x_{3j} + A_3 x_{4j})^T \right) \right] \right\}.$$

Operando de la misma forma se llega a que las distribuciones a posteriori de A_2 y A_3 son de la forma

$$p(A_2 \mid \dots) \propto |A_2|^m \exp \left\{ -\frac{1}{2} \left[\text{tr} \left(A_2 \left(\frac{3}{4\sigma^2} \sum_{j=1}^m x_{3j} x_{3j}^T + I_d \right) A_2^T \right) - \text{tr} \left(\frac{1}{2\sigma^2} A_2 \sum_{j=1}^m x_{3j} (x_{1j} + A_1 x_{2j} + A_3 x_{4j})^T \right) \right] \right\}.$$

$$p(A_3 \mid \dots) \propto |A_3|^m \exp \left\{ -\frac{1}{2} \left[\text{tr} \left(A_3 \left(\frac{3}{4\sigma^2} \sum_{j=1}^m x_{4j} x_{4j}^T + I_d \right) A_3^T \right) - \text{tr} \left(\frac{1}{2\sigma^2} A_3 \sum_{j=1}^m x_{4j} (x_{1j} + A_1 x_{2j} + A_2 x_{3j})^T \right) \right] \right\}.$$

3.6. Modelización con transformaciones euclídeas de similitud

En esta sección presentamos el mismo problema de emparejamiento de $M > 2$ configuraciones de puntos, cuando las transformaciones existentes entre los sistemas de referencia de las configuraciones son transformaciones de similitud. Este tipo de transformaciones multiplican todas las distancias por el mismo factor y son esenciales en el estudio de formas y de emparejamiento de configuraciones. Tal como se explicaba en el capítulo 2, estas transformaciones rotan, trasladan y escalan unas configuraciones sobre otras.

El esquema que se sigue en esta sección y el desarrollo teórico, no son muy diferentes a los presentados anteriormente. En primer lugar se planteará el modelo

general, y se calculará la verosimilitud de los datos. A continuación, se tratará el caso donde se supone normalidad para los errores, además de centrarse en cuatro configuraciones con el objetivo de mostrar aplicaciones. Por último, se realizarán inferencias sobre todos los parámetros suponiendo que los emparejamientos son conocidos.

Así el modelo (3.1) quedaría expresado en términos de matrices de rotación, vectores de traslación y constantes de escalamiento de la siguiente manera:

$$\begin{aligned}
 x_{1j} &= \mu_{\xi_{1j}} + \varepsilon_{1j} & j &= 1, \dots, n_1 \\
 s_1 A_1 x_{2j} + \tau_1 &= \mu_{\xi_{2j}} + \varepsilon_{2j} & j &= 1, \dots, n_2 \\
 &\vdots & &\vdots \\
 s_{r-1} A_{r-1} x_{rj} + \tau_{r-1} &= \mu_{\xi_{rj}} + \varepsilon_{rj} & j &= 1, \dots, n_r \\
 &\vdots & &\vdots \\
 s_{M-1} A_{M-1} x_{Mj} + \tau_{M-1} &= \mu_{\xi_{Mj}} + \varepsilon_{Mj} & j &= 1, \dots, n_M
 \end{aligned} \tag{3.32}$$

donde para todo $i = 1, \dots, M$ y para todo $j = 1, \dots, n_i$, $s_i > 0$ son parámetros de escala, A_i matrices de rotación en \mathbb{R}^d ($|A_i| = 1$ y $A_i^T = A_i^{-1}$), y τ_i vectores de traslación de \mathbb{R}^d . Además, $\{\varepsilon_{ij}\}$ tienen como función de densidad f_i y son independientes entre sí.

Así, a partir de (3.32) se tiene que las funciones de densidad de los distintos puntos de las configuraciones tienen las expresiones,

$$\begin{aligned}
 f(x_{1j}) &= f_1(x_{1j} - \mu_{\xi_{1j}}) & j &= 1, \dots, n_1 \\
 f(x_{2j}) &= f_2(s_1 A_1 x_{2j} + \tau_1 - \mu_{\xi_{2j}}) s_1^d & j &= 1, \dots, n_2 \\
 &\dots & &\dots \\
 f(x_{rj}) &= f_r(s_{r-1} A_{r-1} x_{rj} + \tau_{r-1} - \mu_{\xi_{rj}}) s_{r-1}^d & j &= 1, \dots, n_r \\
 &\dots & &\dots \\
 f(x_{Mj}) &= f_M(s_{M-1} A_{M-1} x_{Mj} + \tau_{M-1} - \mu_{\xi_{Mj}}) s_{M-1}^d & j &= 1, \dots, n_M
 \end{aligned}$$

Por lo tanto, suponiendo $A_0 = I$ y $\tau_0 = 0$, para todo $i = 1, \dots, M$ y para todo $j = 1, \dots, n_i$ la función de densidad de cada punto x_{ij} es,

$$f(x_{ij}) = f_i(s_{i-1} A_{i-1} x_{ij} + \tau_{i-1} - \mu_{\xi_{ij}}) s_{i-1}^d. \tag{3.33}$$

3.6.1. Construcción de la verosimilitud de los datos

De la misma manera que en el caso de transformaciones lineales generales, la verosimilitud de los datos se construye diferenciando los puntos según el grado de emparejamiento que presentan. Las expresiones que se obtienen para cada caso van a ser análogas a las encontradas en aquella situación. Bastará sustituir $|A_k|$ por $|s_k A_k| = s_k^d$ y la expresión $A_{k-1}x_{kj}$ por $s_{k-1}A_{k-1}x_{kj} + \tau_{k-1}$.

Así, a partir de (3.11), la aportación aproximada a la verosimilitud de los puntos no emparejados tiene la expresión,

$$\left(\frac{1}{v}\right)^{n-\sum_{i=2}^M iL_i} \prod_{i=1}^M |s_{i-1}|^{n_i-\sum_{l=2}^M e_i^{(l)}}. \quad (3.34)$$

De la misma manera, a partir de la expresión obtenida en el caso general (3.16), la aportación aproximada a la verosimilitud de los puntos con emparejamiento de orden k , $k = 2, \dots, M$, tiene la forma

$$\begin{aligned} & \left(\frac{1}{v}\right)^{L_k} \prod_{\substack{\{(i_1, \dots, i_k)\} \\ M_{i_1 \dots i_k}^{(k)} \neq 0}} \left(|s_{i_1-1}|^d \cdots |s_{i_k-1}|^d\right)^{M_{i_1 \dots i_k}^{(k)}} \cdot \\ & \prod_{\substack{\{(j_1, \dots, j_k)\} \\ S_{j_1 \dots j_k}^{(i_1 \dots i_k)} = 1}} g_{i_1, i_2, \dots, i_k}(s_{i_1-1}A_{i_1-1}x_{i_1 j_1} + \tau_{i_1-1} - s_{i_2-1}A_{i_2-1}x_{i_2 j_2} - \tau_{i_2-1}, \\ & \dots, s_{i_{k-1}-1}A_{i_{k-1}-1}x_{i_{k-1} j_{k-1}} + \tau_{i_{k-1}-1} - s_{i_k-1}A_{i_k-1}x_{i_k j_k} - \tau_{i_k-1}), \end{aligned} \quad (3.35)$$

donde $g_{i_1, i_2, i_3, \dots, i_k}(z_2, \dots, z_k)$ viene definida en (3.15). El producto de (3.34) y de (3.35) para todo k , da lugar a la expresión aproximada de la verosimilitud en el caso de transformaciones de similitud.

3.6.2. Emparejamientos de $M = 4$ configuraciones con errores normales

Siguiendo con el esquema seguido en la sección 3.5, vamos a suponer que los errores tienen una distribución normal. Así, particularizando en el caso de $M = 4$

configuraciones, el modelo (3.32) queda de la forma

$$\begin{aligned} x_{1j} &= \mu_{\xi_{1j}} + \varepsilon_{1j} & j &= 1, \dots, n_1 \\ s_1 A_1 x_{2j} + \tau_1 &= \mu_{\xi_{2j}} + \varepsilon_{2j} & j &= 1, \dots, n_2 \\ s_2 A_2 x_{3j} + \tau_2 &= \mu_{\xi_{3j}} + \varepsilon_{3j} & j &= 1, \dots, n_3 \\ s_3 A_3 x_{4j} + \tau_3 &= \mu_{\xi_{4j}} + \varepsilon_{4j} & j &= 1, \dots, n_4 \end{aligned}$$

donde $\{\varepsilon_{ij}\}$ son independientes con distribución $\mathbf{N}_d(0, \sigma^2 \mathbf{I}_d)$ con $d \geq 2$ y la matriz \mathbf{I}_d es la matriz identidad de orden d .

En esta situación, los parámetros que definen los emparejamientos (matrices de afinidad y de emparejamientos) son los mismos que en el caso de transformaciones lineales generales, no así los parámetros que definen las transformaciones. Ahora, se tienen, con $i = 1, 2, 3$, las matrices de rotación A_i , los vectores de traslación τ_i y las constantes de escalamiento s_i , además de la varianza de los errores σ^2 .

Las aportaciones a la verosimilitud de los datos de los distintos emparejamientos, se deducen directamente de (3.21), (3.23), (3.24) y (3.25). Así, la aportación de los puntos no emparejados tiene la expresión

$$\left(\frac{1}{v}\right)^{n-2L_2-3L_3-4L_4} \frac{d(n_2 - \sum_{k=2}^M e_2^{(k)})}{s_1} \frac{d(n_3 - \sum_{k=2}^M e_3^{(k)})}{s_2} \frac{d(n_4 - \sum_{k=2}^M e_4^{(k)})}{s_3}. \quad (3.36)$$

La expresión de la aportación de los puntos con emparejamientos dobles es,

$$\begin{aligned} &\left(\frac{1}{v}\right)^{L_2} \prod_{\substack{\{(i_1, i_2)\} \\ M_{i_1, i_2}^{(2)} \neq 0}} (s_{i_1-1}^d s_{i_2-1}^d)^{M_{i_1, i_2}^{(2)}} \prod_{\substack{\{(j_1, j_2)\} \\ S_{j_1 j_2}^{(i_1, i_2)} = 1}} \left(\frac{1}{\sigma\sqrt{2}}\right)^d \\ &\varphi_d\left(\frac{s_{i_1-1} A_{i_1-1} x_{i_1 j_1} + \tau_1 - s_{i_2-1} A_{i_2-1} x_{i_2 j_2} - \tau_2}{\sigma\sqrt{2}}\right). \end{aligned} \quad (3.37)$$

La expresión de los emparejamientos triples es:

$$\begin{aligned} & \left(\frac{1}{v}\right)^{L_3} \prod_{\substack{\{(i_1, i_2, i_3)| \\ M_{i_1, i_2, i_3}^{(3)} \neq 0\}}} (s_{i_1-1} s_{i_2-1} s_{i_3-1})^{dM_{i_1, i_2, i_3}^{(3)}} \cdot \prod_{\substack{\{(j_1, j_2, j_3)| \\ S_{j_1, j_2, j_3}^{(i_1, i_2, i_3)} = 1\}}} \left(\frac{1}{\sigma^2 \sqrt{3}}\right)^d \\ & \cdot \varphi_{2d} \left(\frac{s_{i_1-1} A_{i_1-1} x_{i_1 j_1} + \tau_{i_1-1} - s_{i_2-1} A_{i_2-1} x_{i_2 j_2} - \tau_{i_2-1}}{\sigma \sqrt{2}}, \right. \\ & \left. \frac{\sqrt{6}}{6\sigma} (s_{i_1-1} A_{i_1-1} x_{i_1 j_1} + \tau_{i_1-1} + s_{i_2-1} A_{i_2-1} x_{i_2 j_2} + \tau_{i_2-1} - 2s_{i_3-1} A_{i_3-1} x_{i_3 j_3} - 2\tau_{i_3-1}) \right), \end{aligned} \quad (3.38)$$

y la de los emparejamientos cuádruples es:

$$\begin{aligned} & \left(\frac{1}{v} s_1 s_2 s_3\right)^{dL_4} \prod_{\substack{\{(j_1, j_2, j_3, j_4)| \\ S_{j_1, j_2, j_3, j_4}^{(1, 2, 3)} = 1\}}} \left(\frac{1}{2\sigma^3}\right)^d \varphi_{3d} \left(\frac{x_{1j_1} - s_1 A_1 x_{2j_2} - \tau_1}{\sigma \sqrt{2}}, \right. \\ & \left. \frac{x_{1j_1} + s_1 A_1 x_{2j_2} + \tau_1 - 2s_2 A_2 x_{3j_3} - 2\tau_2}{\sigma \sqrt{6}}, \right. \\ & \left. \frac{x_{1j_1} + s_1 A_1 x_{2j_2} + \tau_1 + s_2 A_2 x_{3j_3} + \tau_2 - 3s_3 A_3 x_{4j_4} - 3\tau_3}{\sigma \sqrt{12}} \right). \end{aligned} \quad (3.39)$$

Por tanto, el producto de (3.36), (3.37), (3.38) y (3.39) da lugar a la expresión de la vesomilitud de los datos de cuatro configuraciones bajo hipótesis de normalidad en el caso de transformaciones de similitud entre las mismas.

3.6.3. Inferencia bayesiana en el caso de emparejamientos conocidos de $M = 4$ configuraciones

En este apartado y de la misma manera que se hacía en el apartado 3.5.3, se presenta el caso de $M = 4$ configuraciones con el mismo número de puntos, m , suponiendo que estos m puntos están emparejados y se conoce cuáles son estos emparejamientos. Así, centrándonos en el caso de que los errores siguen una distribución normal, los parámetros del modelo se reducen a aquellos con los que se definen las transformaciones entre configuraciones, es decir, los parámetros $A_1, A_2, A_3, s_1, s_2, s_3, \tau_1, \tau_2, \tau_3$ y el parámetro de varianza σ^2 .

Supongamos sin pérdida de generalidad que $\forall j = 1, \dots, m$ los puntos x_{1j}, x_{2j}, x_{3j} y x_{4j} forman los m emparejamientos de orden cuatro, es decir, $\xi_{1j} = \xi_{2j} = \xi_{3j} = \xi_{4j}$. En esta situación, las matrices de afinidad y de emparejamientos son conocidas y son las descritas en el apartado 3.5.3.

Basándose en las expresiones (3.36), (3.37), (3.38) y (3.39) se obtiene la expresión de la verosimilitud, análoga a (3.26),

$$p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \mid A_1, A_2, A_3, s_1, s_2, s_3, \tau_1, \tau_2, \tau_3, \sigma^2) \propto \left(\frac{1}{\sigma^2} \right)^{\frac{3}{2}md} (s_1 s_2 s_3)^{md} \cdot \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{j=1}^m \|D_{1j}\|^2 + \sum_{j=1}^m \|D_{2j}\|^2 + \sum_{j=1}^m \|D_{3j}\|^2 \right) \right\}, \quad (3.40)$$

donde

$$\begin{aligned} D_{1j} &= \frac{1}{\sqrt{2}}(x_{1j} - s_1 A_1 x_{2j} - \tau_1), \\ D_{2j} &= \frac{1}{\sqrt{6}}(x_{1j} + s_1 A_1 x_{2j} + \tau_1 - 2s_2 A_2 x_{3j} - 2\tau_2), \\ D_{3j} &= \frac{1}{\sqrt{12}}(x_{1j} + s_1 A_1 x_{2j} + \tau_1 + s_2 A_2 x_{3j} + \tau_2 - 3s_3 A_3 x_{4j} - 3\tau_3). \end{aligned} \quad (3.41)$$

Distribución a priori de la varianza σ^2 y su distribución a posteriori

Partiendo de que la distribución a priori de la varianza σ^2 es una gamma inversa $\sigma^{-2} \sim G(\alpha, \beta)$, se puede deducir de la misma manera que en el caso de transformaciones generales, que la distribución a posteriori de σ^2 sigue siendo una gamma inversa $\sigma^{-2} \sim G(\alpha^*, \beta^*)$ de parámetros,

$$\begin{aligned} \alpha^* &= \alpha + \frac{3}{2}md \\ \beta^* &= \beta + \left\{ \frac{1}{2} \sum_{j=1}^m \|x_{1j} - s_1 A_1 x_{2j} - \tau_1\|^2 + \frac{1}{6} \sum_{j=1}^m \|x_{1j} + s_1 A_1 x_{2j} + \tau_1 - 2s_2 A_2 x_{3j} - 2\tau_2\|^2 \right. \\ &\quad \left. + \frac{1}{12} \sum_{j=1}^m \|x_{1j} + s_1 A_1 x_{2j} + \tau_1 + s_2 A_2 x_{3j} + \tau_2 - 3s_3 A_3 x_{4j} - 3\tau_3\|^2 \right\}, \end{aligned}$$

expresiones análogas a (3.27), obtenidas en aquel caso.

Distribuciones a priori de los parámetros de traslación τ_1, τ_2 y τ_3 y sus distribuciones a posteriori

Asumimos que la distribución a priori de los vectores de traslación es normal, $\tau_i \sim N_d(\mu_i, \sigma_{\tau_i}^2 \mathbf{I}_d)$ $i = 1, 2, 3$ es decir,

$$p(\tau_i) \propto \exp \left\{ -\frac{1}{2\sigma_{\tau_i}^2} (\tau_i - \mu_i)^T (\tau_i - \mu_i) \right\}.$$

Bajo el modelo (3.40) se tiene que la distribución a posteriori de τ_1 será de la forma,

$$\begin{aligned} p(\tau_1 | \dots) &\propto \exp \left\{ -\frac{1}{2\sigma_{\tau_1}^2} (\tau_1 - \mu_1)^T (\tau_1 - \mu_1) \right\} \cdot \\ &\exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{j=1}^m \|D_{1j}\|^2 + \sum_{j=1}^m \|D_{2j}\|^2 + \sum_{j=1}^m \|D_{3j}\|^2 \right) \right\} \\ &= \exp \left\{ -\frac{1}{2\sigma_1^2} (\tau_1 - \mu_1)^T (\tau_1 - \mu_1) - \frac{1}{2\sigma^2} \left(\sum_{j=1}^m \|D_{1j}\|^2 \right. \right. \\ &\quad \left. \left. + \sum_{j=1}^m \|D_{2j}\|^2 + \sum_{j=1}^m \|D_{3j}\|^2 \right) \right\}. \end{aligned} \quad (3.42)$$

Operamos en (3.42) con el objetivo de llegar a una expresión simplificada.

$$\begin{aligned} \sum_{j=1}^m \|D_{1j}\|^2 &= \sum_{j=1}^m \left\| \frac{1}{\sqrt{2}} (x_{1j} - s_1 A_1 x_{2j} - \tau_1) \right\|^2 \\ &= \frac{1}{2} (m\tau_1^T \tau_1 - 2\tau_1^T \sum_{j=1}^m (x_{1j} - s_1 A_1 x_{2j})) + K_1, \end{aligned}$$

$$\begin{aligned} \sum_{j=1}^m \|D_{2j}\|^2 &= \sum_{j=1}^m \left\| \frac{1}{\sqrt{6}} (x_{1j} + s_1 A_1 x_{2j} + \tau_1 - 2s_2 A_2 x_{3j} - 2\tau_2) \right\|^2 \\ &= \frac{1}{6} (m\tau_1^T \tau_1 - 2\tau_1^T \sum_{j=1}^m (2s_2 A_2 x_{3j} + 2\tau_2 - x_{1j} - s_1 A_1 x_{2j})) + K_2, \end{aligned}$$

$$\begin{aligned} \sum_{j=1}^m \|D_{3j}\|^2 &= \sum_{j=1}^m \left\| \frac{1}{\sqrt{12}} (x_{1j} + s_1 A_1 x_{2j} + \tau_1 + s_2 A_2 x_{3j} + \tau_2 - 3s_3 A_3 x_{4j} - 3\tau_3) \right\|^2 \\ &= \frac{1}{12} (m\tau_1^T \tau_1 - 2\tau_1^T \sum_{j=1}^m (3s_3 A_3 x_{4j} + 3\tau_3 - x_{1j} - s_1 A_1 x_{2j} - s_2 A_2 x_{3j} - \tau_2)) \\ &\quad + K_3, \end{aligned}$$

con K_1, K_2 y K_3 expresiones independientes de τ_1 .

Así, el exponente en (3.42) sería,

$$\begin{aligned}
 & -\frac{1}{2\sigma_{\tau_1}^2}(\tau_1^T \tau_1 - 2\tau_1^T \mu_{\tau_1}) - \frac{1}{2\sigma^2} \left(\frac{1}{2}(m\tau_1^T \tau_1 - 2\tau_1^T \sum_{j=1}^m (x_{1j} - s_1 A_1 x_{2j})) \right. \\
 & + \frac{1}{6}(m\tau_1^T \tau_1 - 2\tau_1^T \sum_{j=1}^m (2s_2 A_2 x_{3j} + 2\tau_2 - x_{1j} - s_1 A_1 x_{2j})) \\
 & \left. + \frac{1}{12}(m\tau_1^T \tau_1 - 2\tau_1^T \sum_{j=1}^m (3s_3 A_3 x_{4j} + 3\tau_3 - x_{1j} - s_1 A_1 x_{2j} - s_2 A_2 x_{3j} - \tau_2)) \right) \\
 = & -\frac{1}{2\sigma_{\tau_1}^2}(\tau_1^T \tau_1 - 2\tau_1^T \mu_{\tau_1}) - \frac{1}{2\sigma^2} \left(\frac{1}{2}(m\tau_1^T \tau_1 - 2\tau_1^T \sum_{j=1}^m C_{1j}) \right. \\
 & \left. + \frac{1}{6}(m\tau_1^T \tau_1 - 2\tau_1^T \sum_{j=1}^m C_{2j}) + \frac{1}{12}(m\tau_1^T \tau_1 - 2\tau_1^T \sum_{j=1}^m C_{3j}) \right),
 \end{aligned}$$

donde

$$\begin{aligned}
 C_{1j} &= x_{1j} - s_1 A_1 x_{2j}, \\
 C_{2j} &= 2s_2 A_2 x_{3j} + 2\tau_2 - x_{1j} - s_1 A_1 x_{2j} \\
 C_{3j} &= 3s_3 A_3 x_{4j} + 3\tau_3 - x_{1j} - s_1 A_1 x_{2j} - s_2 A_2 x_{3j} - \tau_2.
 \end{aligned}$$

Entonces, la densidad a posteriori de τ_1 será de la forma:

$$\begin{aligned}
 p(\tau_1 \mid \dots) &\propto \\
 &\exp \left\{ -\frac{1}{2} \left[\frac{1}{12\sigma_{\tau_1}^2 \sigma^2} (\tau_1^T \tau_1 (12\sigma^2 + 9m\sigma_{\tau_1}^2) \right. \right. \\
 &\quad \left. \left. - 2\tau_1^T \left(12\mu_{\tau_1} \sigma^2 + 6\sigma_{\tau_1}^2 \sum_{j=1}^m C_{1j} + 2\sigma_{\tau_1}^2 \sum_{j=1}^m C_{2j} + \sigma_{\tau_1}^2 \sum_{j=1}^m C_{3j} \right) \right] \right\}.
 \end{aligned}$$

Completando cuadrados con el fin de obtener la expresión de la función de den-

sidad de una distribución normal y operando, se llega a que,

$$p(\tau_1 | \dots) \propto \exp \left\{ -\frac{1}{2 \frac{\sigma_{\tau_1}^2 \sigma^2}{\sigma^2 + \frac{3}{4} m \sigma_{\tau_1}^2}} \left\| \tau_1 - \frac{\frac{\mu_{\tau_1}}{\sigma_{\tau_1}^2} + \frac{1}{4\sigma^2} \sum_{j=1}^m (x_{1j} - 3s_1 A_1 x_{2j} + s_2 A_2 x_{3j} + s_3 A_3 x_{4j} + \tau_2 + \tau_3)}{\frac{1}{\sigma_{\tau_1}^2} + \frac{3m}{4\sigma^2}} \right\|^2 \right\}.$$

Por lo tanto, la distribución a posteriori de τ_1 es normal d -dimensional

$$\tau_1 | \dots \sim N_d \left(\frac{\frac{\mu_{\tau_1}}{\sigma_{\tau_1}^2} + \frac{1}{4\sigma^2} \sum_{j=1}^m (x_{1j} - 3s_1 A_1 x_{2j} + s_2 A_2 x_{3j} + s_3 A_3 x_{4j} + \tau_2 + \tau_3)}{\frac{1}{\sigma_{\tau_1}^2} + \frac{3m}{4\sigma^2}}, \frac{1}{\frac{1}{\sigma_{\tau_1}^2} + \frac{3m}{4\sigma^2}} \mathbf{I}_d \right).$$

Actuando de la misma manera para el cálculo de la distribución a posteriori de τ_2 y operando se llega a que,

$$p(\tau_2 | \dots) \propto \exp \left\{ -\frac{1}{2 \left(1 / \left(\frac{1}{\sigma_{\tau_2}^2} + \frac{3m}{4\sigma^2} \right) \right)} \left\| \tau_2 - \frac{\frac{\mu_{\tau_2}}{\sigma_{\tau_2}^2} + \frac{1}{4\sigma^2} \sum_{j=1}^m (x_{1j} + s_1 A_1 x_{2j} - 3s_2 A_2 x_{3j} + s_3 A_3 x_{4j} + \tau_1 + \tau_3)}{\frac{1}{\sigma_{\tau_2}^2} + \frac{3m}{4\sigma^2}} \right\|^2 \right\},$$

por lo que la distribución a posteriori de τ_2 es también normal d -dimensional ,

$$\tau_2 | \dots \sim N_d \left(\frac{\frac{\mu_{\tau_2}}{\sigma_{\tau_2}^2} + \frac{1}{4\sigma^2} \sum_{j=1}^m (x_{1j} + s_1 A_1 x_{2j} - 3s_2 A_2 x_{3j} + s_3 A_3 x_{4j} + \tau_1 + \tau_3)}{\frac{1}{\sigma_{\tau_2}^2} + \frac{3m}{4\sigma^2}}, \frac{1}{\frac{1}{\sigma_{\tau_2}^2} + \frac{3m}{4\sigma^2}} \mathbf{I}_d \right). \quad (3.43)$$

Por último, y de la misma manera, se llega a que la distribución a posteriori de τ_3 es,

$$\tau_3 | \dots \sim N_d \left(\frac{\frac{\mu_{\tau_3}}{\sigma_{\tau_3}^2} + \frac{1}{4\sigma^2} \sum_{j=1}^m (x_{1j} + s_1 A_1 x_{2j} + s_2 A_2 x_{3j} - 3s_3 A_3 x_{4j} + \tau_1 + \tau_2)}{\frac{1}{\sigma_{\tau_3}^2} + \frac{3m}{4\sigma^2}}, \frac{1}{\frac{1}{\sigma_{\tau_3}^2} + \frac{3m}{4\sigma^2}} \mathbf{I}_d \right). \quad (3.44)$$

Distribuciones a priori de los parámetros de escala s_1, s_2 y s_3 y sus distribuciones a posteriori

Asumimos que a priori, los parámetros de escala s_i , $i = 1, 2, 3$, siguen una distribución gamma, es decir $s_i \sim G(\alpha_i, \beta_i)$, cuya función de densidad es,

$$p(s_i) = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} s_i^{\alpha_i-1} \exp\{-\beta_i s_i\} \quad s_i > 0.$$

Entonces a partir de (3.40), la distribución a posteriori de s_1 tendrá la forma:

$$p(s_1 | \dots) \propto s_1^{\alpha-1} \exp\{-\beta_1 s_1\} s_1^{md} \exp\left\{-\frac{1}{2\sigma^2} \left(\sum_{j=1}^m \|D_{1j}\|^2 + \sum_{j=1}^m \|D_{2j}\|^2 + \sum_{j=1}^m \|D_{3j}\|^2\right)\right\},$$

donde D_{1j}, D_{2j} y D_{3j} son como en (3.41).

Desarrollando las normas al cuadrado se llega a que

$$\begin{aligned} \sum_{j=1}^m \|D_{1j}\|^2 + \|D_{2j}\|^2 + \|D_{3j}\|^2 &= s_1^2 \frac{3}{4} \sum_{j=1}^m \|x_{2j}\|^2 \\ &- \frac{1}{2} s_1 \sum_{j=1}^m (x_{1j} - 3\tau_1 + s_2 A_2 x_{3j} + \tau_2 + s_3 A_3 x_{4j} + \tau_3)^T A_1 x_{2j} + K_1, \end{aligned}$$

donde K_1 es una expresión independiente de s_1 .

Así, la función de densidad a posteriori del parámetro s_1 es:

$$\begin{aligned} p(s_1 | \dots) &\propto s_1^{\alpha+dm-1} \exp\left\{-\frac{1}{2\sigma^2} \left(s_1^2 \frac{3}{4} \sum_{j=1}^m \|x_{2j}\|^2 \right. \right. \\ &\quad \left. \left. - s_1 \left[\frac{1}{2} \sum_{j=1}^m (x_{1j} - 3\tau_1 + s_2 A_2 x_{3j} + \tau_2 + s_3 A_3 x_{4j} + \tau_3)^T A_1 x_{2j} - 2\sigma^2 \beta_1 \right] \right) \right\}. \end{aligned}$$

Procediendo de la misma manera para el cálculo de la distribución a posteriori de s_2 , y observando que D_{1j} no depende de s_2 , se tiene que:

$$p(s_2 | \dots) \propto s_2^{\alpha-1} \exp\{-\beta_2 s_2\} s_2^{md} \exp\left\{-\frac{1}{2\sigma^2} \left(\sum_{j=1}^m \|D_{2j}\|^2 + \sum_{j=1}^m \|D_{3j}\|^2\right)\right\}.$$

Operando de nuevo, se llega a que

$$\begin{aligned} \|D_{2j}\|^2 + \|D_{3j}\|^2 &= s_2^2 \frac{3}{4} \sum_{j=1}^m \|x_{3j}\|^2 \\ &- \frac{1}{2} s_2 \sum_{j=1}^m (x_{1j} + \tau_1 + s_1 A_1 x_{2j} - 3\tau_2 + s_3 A_3 x_{4j} + \tau_3)^T A_2 x_{3j} + K_2, \end{aligned}$$

donde K_2 es una expresión independiente de s_2 . Así, la función de densidad a posteriori del parámetro s_2 es:

$$\begin{aligned} p(s_2 \mid \dots) &\propto s_2^{\alpha+dm-1} \exp \left\{ -\frac{1}{2\sigma^2} \left(s_2^2 \frac{3}{4} \sum_{j=1}^m \|x_{3j}\|^2 \right. \right. \\ &\quad \left. \left. - s_2 \left[\frac{1}{2} \sum_{j=1}^m (x_{1j} + \tau_1 + s_1 A_1 x_{2j} - 3\tau_2 + s_3 A_3 x_{4j} + \tau_3)^T A_2 x_{3j} - 2\sigma^2 \beta_2 \right] \right) \right\}. \end{aligned}$$

Por último, en el caso del parámetro s_3 , se observa que D_{1j} y D_{2j} no dependen de s_3 , por lo que

$$p(s_3 \mid \dots) \propto s_3^{\alpha-1} \exp\{-\beta_3 s_3\} s_3^{md} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^m \|D_{3j}\|^2 \right\}.$$

En este caso, y de forma análoga, se llega a que la función de densidad a posteriori de s_3 es:

$$\begin{aligned} p(s_3 \mid \dots) &\propto s_3^{\alpha+dm-1} \exp \left\{ -\frac{1}{2\sigma^2} \left(s_3^2 \frac{3}{4} \sum_{j=1}^m \|x_{4j}\|^2 \right. \right. \\ &\quad \left. \left. - s_3 \left[\frac{1}{2} \sum_{j=1}^m (x_{1j} + \tau_1 + s_1 A_1 x_{2j} + \tau_2 + s_2 A_2 x_{3j} - 3\tau_3)^T A_3 x_{4j} - 2\sigma^2 \beta_3 \right] \right) \right\}. \end{aligned}$$

Distribuciones a priori de las matrices de rotación A_1, A_2 y A_3 y sus distribuciones a posteriori

A partir de este momento, nos situamos en \mathbb{R}^2 con el fin de centrarnos en matrices de rotación en dicho espacio. Elijiendo como sentido de la rotación el contrario a las agujas del reloj, las matrices A_1, A_2 y A_3 serán de la forma,

$$A_1 = \begin{pmatrix} \cos \theta_1 & -\text{sen} \theta_1 \\ \text{sen} \theta_1 & \cos \theta_1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} \cos \theta_2 & -\text{sen} \theta_2 \\ \text{sen} \theta_2 & \cos \theta_2 \end{pmatrix}, \quad A_3 = \begin{pmatrix} \cos \theta_3 & -\text{sen} \theta_3 \\ \text{sen} \theta_3 & \cos \theta_3 \end{pmatrix}.$$

Siguiendo a Green y Mardia (2006), asumimos que las distribuciones a priori de las matrices de rotación A_1, A_2 y A_3 son distribuciones von Mises de parámetros ν_i y $k_i > 0$ (Mardia y Jupp, 2000) con $i = 1, 2, 3$, $A_i \sim M(\nu_i, k_i)$, es decir, sus funciones de densidad son,

$$\begin{aligned} p(A_i) &= \frac{1}{2\pi I_0(k_i)} \exp \{ \text{traza} (F_i^T A_i) \} \\ &\propto \exp \{ \text{traza} (F_i^T A_i) \}, \end{aligned}$$

donde I_0 denota la función de Bessel modificada, definida como

$$I_0(k_i) = \frac{1}{2\pi} \int_0^{2\pi} \exp(k_i \cos \theta) d\theta,$$

y la matriz

$$F_i = \frac{k_i}{2} \begin{pmatrix} \cos \nu_i & -\text{sen} \nu_i \\ \text{sen} \nu_i & \cos \nu_i \end{pmatrix}.$$

Entonces, las funciones de densidad se pueden escribir también en términos del ángulo θ_i , como

$$p(\theta_i) \propto \exp \{ k_i \cos \nu_i \cos \theta_i + k_i \text{sen} \nu_i \text{sen} \theta_i \} = \exp \{ k_i \cos(\theta_i - \nu_i) \}.$$

Para calcular la distribución a posteriori de A_1 , nos basamos en el modelo (3.40). Así,

$$\begin{aligned} p(A_1 \mid \dots) &\propto \exp \{ \text{traza} (F_1^T A_1) \} \cdot \\ &\exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{j=1}^m \|D_{1j}\|^2 + \sum_{j=1}^m \|D_{2j}\|^2 + \sum_{j=1}^m \|D_{3j}\|^2 \right) \right\}. \end{aligned} \quad (3.45)$$

Operando y prescindiendo de lo que depende de A_1 , se tiene que,

$$\begin{aligned} \|D_{1j}\|^2 &= \left\| \frac{1}{\sqrt{2}} (x_{1j} - s_1 A_1 x_{2j} - \tau_1) \right\|^2 \\ &= \frac{1}{2} (\|A_1 x_{2j}\|^2 - 2s_1 (x_{1j} - \tau_1)^T A_1 x_{2j} + \|x_{1j} - \tau_1\|^2) \\ &= -s_1 (x_{1j} - \tau_1)^T A_1 x_{2j} + K_1, \end{aligned}$$

$$\begin{aligned}
 \|D_{2j}\|^2 &= \left\| \frac{1}{\sqrt{6}}(x_{1j} + s_1 A_1 x_{2j} + \tau_1 - 2s_2 A_2 x_{3j} - 2\tau_2) \right\|^2 \\
 &= \frac{1}{6} (\|s_1 A_1 x_{2j}\|^2 - 2s_1 (2s_2 A_2 x_{3j} + 2\tau_2 - x_{1j} - \tau_1)^T A_1 x_{2j} \\
 &\quad + \|2s_2 A_2 x_{3j} + 2\tau_2 - x_{1j} - \tau_1\|^2) \\
 &= -\frac{1}{3} s_1 (2s_2 A_2 x_{3j} + 2\tau_2 - x_{1j} - \tau_1)^T A_1 x_{2j},
 \end{aligned}$$

y

$$\begin{aligned}
 \|D_{3j}\|^2 &= \left\| \frac{1}{\sqrt{12}}(x_{1j} + s_1 A_1 x_{2j} + \tau_1 + s_2 A_2 x_{3j} + \tau_2 - 3s_3 A_3 x_{4j} - 3\tau_3) \right\|^2 \\
 &= -\frac{1}{6} s_1 (3s_3 A_3 x_{4j} + 3\tau_3 - x_{1j} - \tau_1 - s_2 A_2 x_{3j} - \tau_2)^T A_1 x_{2j} + K_3,
 \end{aligned}$$

donde K_i son expresiones independientes de A_1 .

Así,

$$\begin{aligned}
 \|D_{1j}\|^2 + \|D_{2j}\|^2 + \|D_{3j}\|^2 &= \\
 -\frac{1}{2} s_1 (x_{1j} + s_2 A_2 x_{3j} + s_3 A_3 x_{4j} - 3\tau_1 + \tau_2 + \tau_3)^T A_1 x_{2j} + K_1 + K_2 + K_3.
 \end{aligned} \tag{3.46}$$

Sustituyendo (3.46) en (3.45) y de nuevo utilizando que $a^T b = \text{traza}(ba^T)$, se llega a que

$$\begin{aligned}
 p(A_1 \mid \dots) &\propto \exp \{ \text{traza} (F_1^T A_1) \} \\
 &\propto \exp \left\{ \text{traza} \left(\frac{1}{4\sigma^2} s_1 \sum_{j=1}^m x_{2j} (x_{1j} + s_2 A_2 x_{3j} + s_3 A_3 x_{4j} - 3\tau_1 + \tau_2 + \tau_3)^T A_1 \right) \right\} \\
 &\propto \exp \left\{ \text{traza} \left(\left[F_1^T + \frac{1}{4\sigma^2} s_1 \sum_{j=1}^m x_{2j} (x_{1j} + s_2 A_2 x_{3j} + s_3 A_3 x_{4j} - 3\tau_1 + \tau_2 + \tau_3)^T \right] A_1 \right) \right\} \\
 &\propto \exp \left\{ \text{traza} \left(\left[F_1 + \frac{1}{4\sigma^2} s_1 \sum_{j=1}^m (x_{1j} + s_2 A_2 x_{3j} + s_3 A_3 x_{4j} - 3\tau_1 + \tau_2 + \tau_3) x_{2j}^T \right]^T A_1 \right) \right\}.
 \end{aligned}$$

Por lo tanto, la distribución a posteriori de A_1 es una distribución von Mises con función de densidad

$$p(A_1 \mid \dots) \propto \exp \{ \text{traza} (F_1^{*T} A_1) \}, \tag{3.47}$$

donde la matriz F_1^* toma la expresión

$$F_1^* = F_1 + \frac{1}{4\sigma^2} s_1 \sum_{j=1}^m (x_{1j} + s_2 A_2 x_{3j} + s_3 A_3 x_{4j} - 3\tau_1 + \tau_2 + \tau_3) x_{2j}^T.$$

Actuando de la misma manera para el cálculo de la distribución a posteriori de A_2 se llega a que,

$$p(A_2 | \dots) \propto \exp \left\{ \text{traza} \left(\left[F_2 + \frac{1}{4\sigma^2} s_2 \sum_{j=1}^m (x_{1j} + s_1 A_1 x_{2j} + s_3 A_3 x_{4j} + \tau_1 - 3\tau_2 + \tau_3) x_{3j}^T \right]^T A_2 \right) \right\},$$

por lo que la distribución a posteriori de A_2 es una distribución von Mises con función de densidad

$$p(A_2 | \dots) \propto \exp \left\{ \text{traza} (F_2^{*T} A_2) \right\}, \quad (3.48)$$

donde la matriz F_2^* toma la expresión

$$F_2^* = F_2 + \frac{1}{4\sigma^2} s_2 \sum_{j=1}^m (x_{1j} + s_1 A_1 x_{2j} + s_3 A_3 x_{4j} + \tau_1 - 3\tau_2 + \tau_3) x_{3j}^T.$$

Por último, se prueba que la distribución a posteriori de A_3 es una distribución von Mises con función de densidad

$$p(A_3 | \dots) \propto \exp \left\{ \text{traza} (F_3^{*T} A_3) \right\}, \quad (3.49)$$

donde la matriz F_3^* toma la expresión

$$F_3^* = F_3 + \frac{1}{4\sigma^2} s_3 \sum_{j=1}^m (x_{1j} + s_1 A_1 x_{2j} + s_2 A_2 x_{3j} + \tau_1 + \tau_2 - 3\tau_3) x_{4j}^T.$$

Con la intención de identificar los parámetros de estas distribuciones von Mises, llamamos

$$T^{(1)} = \begin{pmatrix} T_{11}^{(1)} & T_{12}^{(1)} \\ T_{21}^{(1)} & T_{22}^{(1)} \end{pmatrix} = \frac{s_1}{4\sigma^2} \sum_{j=1}^m (x_{1j} + s_2 A_2 x_{3j} + s_3 A_3 x_{4j} - 3\tau_1 + \tau_2 + \tau_3) x_{2j}^T,$$

$$T^{(2)} = \begin{pmatrix} T_{11}^{(2)} & T_{12}^{(2)} \\ T_{21}^{(2)} & T_{22}^{(2)} \end{pmatrix} = \frac{s_2}{4\sigma^2} \sum_{j=1}^m (x_{1j} + s_1 A_1 x_{2j} + s_3 A_3 x_{4j} + \tau_1 - 3\tau_2 + \tau_3) x_{3j}^T,$$

y

$$T^{(3)} = \begin{pmatrix} T_{11}^{(3)} & T_{12}^{(3)} \\ T_{21}^{(3)} & T_{22}^{(3)} \end{pmatrix} = \frac{s_3}{4\sigma^2} \sum_{j=1}^m (x_{1j} + s_1 A_1 x_{2j} + s_2 A_2 x_{3j} + \tau_1 + \tau_2 - 3\tau_3) x_{4j}^T,$$

por lo que $F_i^* = F_i + T^{(i)}$ para todo $i = 1, 2, 3$. Entonces, las expresiones (3.47), (3.48) y (3.49) pueden escribirse como

$$p(A_i | \dots) \propto \exp \left\{ \text{traza} \left((F_i + T^{(i)})^T A_i \right) \right\},$$

o también

$$p(\theta_i | \dots) \propto \exp \left\{ (k_i \cos \nu_i + T_{11}^{(i)} + T_{22}^{(i)}) \cos \theta_i + (k_i \sin \nu_i + T_{21}^{(i)} - T_{12}^{(i)}) \sin \theta_i \right\}.$$

Llamando ν_i^* y $k_i^* > 0$ a los parámetros de la distribución von Mises a posteriori de A_i y resolviendo el sistema de ecuaciones

$$\begin{aligned} k_i \cos \nu_i + T_{11}^{(i)} + T_{22}^{(i)} &= k_i^* \cos \nu_i^*, \\ k_i \sin \nu_i + T_{21}^{(i)} - T_{12}^{(i)} &= k_i^* \sin \nu_i^*, \end{aligned}$$

se llega a que para todo $i = 1, 2, 3$, la distribución a posteriori de A_i es von Mises de parámetros $A_i \sim M(\nu_i^*, k_i^*)$ donde

$$\begin{aligned} k_i^* &= \left[\left(k_i \cos \nu_i + T_{11}^{(i)} + T_{22}^{(i)} \right)^2 + \left(k_i \sin \nu_i + T_{21}^{(i)} - T_{12}^{(i)} \right)^2 \right]^{1/2}, \\ \nu_i^* &= \arccos \left[\frac{k_i \cos \nu_i + T_{11}^{(i)} + T_{22}^{(i)}}{k_i^*} \right]. \end{aligned}$$

3.6.4. Inferencia bayesiana en el caso de emparejamientos desconocidos de $M = 3$ configuraciones y transformaciones rígidas

Hasta ahora, las inferencias de los parámetros que se han planteado no involucraban ninguna matriz de emparejamientos ni de similitud al trabajarse en el supuesto de conocerse los emparejamientos entre las configuraciones. En este apartado vamos a estudiar el caso en el que las configuraciones siguen teniendo el mismo número de puntos m , se sabe que tienen todos sus puntos emparejados, pero se desconoce cuáles son estos emparejamientos. Esta situación la vamos a desarrollar en el caso de $M = 3$ configuraciones para simplificar, y vamos a utilizar transformaciones rígidas, es decir, suponemos los parámetros de escalamiento $s_i = 1$ y mantenemos la hipótesis de normalidad sobre los errores.

Así, en esta situación, los parámetros del modelo son las matrices de rotación A_1, A_2 , los vectores de traslación τ_1, τ_2 , el parámetro de precisión σ^2 y la matriz de emparejamientos triples $S^{(1,2,3)}$. Señalamos que la matriz de afinidad $\mathbf{M}^{(3)}$ es conocida, ya que sus elementos son todos cero excepto $M_{123}^{(3)} = m$.

Comenzamos determinando la verosimilitud de los datos. A partir de (3.38) y particularizando en el caso de tres configuraciones sin parámetros de escala, esta verosimilitud tiene la forma:

$$\begin{aligned}
 p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \mid A_1, A_2, \tau_1, \tau_2, \sigma^2, S^{(1,2,3)}) &\approx \left(\frac{1}{v}\right)^m \prod_{\substack{\{(i,j,k)\} \\ S_{i,j,k}^{(1,2,3)}=1}} \left(\frac{1}{\sigma^2\sqrt{3}}\right)^d \cdot \\
 &\varphi_{2d} \left(\frac{x_{1i} - A_1x_{2j} - \tau_1}{\sigma\sqrt{2}}, \frac{\sqrt{6}}{6\sigma}(x_{1i} + A_1x_{2j} + \tau_1 - 2A_2x_{3k} - 2\tau_2) \right) \\
 &\propto \left(\frac{1}{\sigma^2\sqrt{3}}\right)^{md} \exp \left\{ -\frac{1}{2} \sum_{\substack{\{(i,j,k)\} \\ S_{i,j,k}^{(1,2,3)}=1}} \left\| \frac{x_{1i} - A_1x_{2j} - \tau_1}{\sigma\sqrt{2}} \right\|^2 \right. \\
 &\quad \left. + \frac{\sqrt{6}}{6\sigma} (x_{1i} + A_1x_{2j} + \tau_1 - 2A_2x_{3k} - 2\tau_2)^2 \right\} \\
 &\propto \left(\frac{1}{\sigma^2\sqrt{3}}\right)^{md} \exp \left\{ -\frac{1}{\sigma^2} \left(\frac{1}{4} \sum_{\substack{\{(i,j,k)\} \\ S_{i,j,k}^{(1,2,3)}=1}} \|x_{1i} - A_1x_{2j} - \tau_1\|^2 + \right. \right. \\
 &\quad \left. \left. + \frac{1}{3} \sum_{\substack{\{(i,j,k)\} \\ S_{i,j,k}^{(1,2,3)}=1}} \left\| \frac{x_{1i} + A_1x_{2j} + \tau_1}{2} - A_2x_{3k} - \tau_2 \right\|^2 \right) \right\}. \quad (3.50)
 \end{aligned}$$

Así, el modelo conjunto tendrá la forma

$$\begin{aligned}
 p(A_1, A_2, \tau_1, \tau_2, \sigma^2, S^{(1,2,3)}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) &\propto p(A_1)p(A_2)p(\tau_1)p(\tau_2)p(S^{(1,2,3)}) \cdot \\
 &\left(\frac{1}{\sigma^2\sqrt{3}}\right)^{md} \exp \left\{ -\frac{1}{\sigma^2} \left(\frac{1}{4} \sum_{\substack{\{(i,j,k)\} \\ S_{i,j,k}^{(1,2,3)}=1}} \|x_{1i} - A_1x_{2j} - \tau_1\|^2 \right. \right. \\
 &\quad \left. \left. + \frac{1}{3} \sum_{\substack{\{(i,j,k)\} \\ S_{i,j,k}^{(1,2,3)}=1}} \left\| \frac{x_{1i} + A_1x_{2j} + \tau_1}{2} - A_2x_{3k} - \tau_2 \right\|^2 \right) \right\}. \quad (3.51)
 \end{aligned}$$

Distribución a priori de la varianza σ^2 y su distribución a posteriori

Si partimos de que la distribución a priori de la varianza σ^2 es la gamma inversa $\sigma^{-2} \sim G(\alpha, \beta)$, a partir de (3.50) se demuestra de la misma forma que en el apartado 3.6.3, que la distribución a posteriori de σ^2 es una gamma inversa $\sigma^{-2} \sim G(\alpha^*, \beta^*)$ donde

$$\begin{aligned} \alpha^* &= \alpha + md \\ \beta^* &= \beta + \left\{ \frac{1}{4} \sum_{\substack{\{(i,j,k)| \\ S_{i,j,k}^{(1,2,3)}=1\}}}^m \|x_{1i} - A_1 x_{2j} - \tau_1\|^2 \right. \\ &\quad \left. + \frac{1}{3} \sum_{\substack{\{(i,j,k)| \\ S_{i,j,k}^{(1,2,3)}=1\}}}^m \|x_{1i} + A_1 x_{2j} + \tau_1 - 2A_2 x_{3k} - 2\tau_2\|^2 \right\}. \end{aligned}$$

Distribuciones a priori de los parámetros de traslación τ_1 y τ_2 y sus distribuciones a posteriori

Asumimos que $\tau_i \sim N_d(\mu_i, \sigma_{\tau_i}^2 \mathbf{I}_d)$ $i = 1, 2$, es decir,

$$p(\tau_i) \propto \exp \left\{ -\frac{1}{2\sigma_{\tau_i}^2} (\tau_i - \mu_i)' (\tau_i - \mu_i) \right\}.$$

Bajo el modelo (3.50) y de la misma forma con la que se procedía en el apartado (3.6.3), las distribuciones a posteriori de τ_1 y τ_2 son, respectivamente

$$\begin{aligned} \tau_1 \mid \dots &\sim \mathbf{N}_d \left(\frac{\frac{\mu_1}{\sigma_{\tau_1}^2} + \frac{1}{3\sigma^2} \sum_{\substack{\{(i,j,k) \mid \\ S_{i,j,k}^{(1,2,3)}=1\}}} (x_{1i} - 2A_1x_{2j} + A_2x_{3k} + \tau_2)}{\frac{1}{\sigma_{\tau_1}^2} + \frac{2m}{3\sigma^2}}, \frac{1}{\frac{1}{\sigma_{\tau_1}^2} + \frac{2m}{3\sigma^2}} \mathbf{I}_d \right), \\ \tau_2 \mid \dots &\sim \mathbf{N}_d \left(\frac{\frac{\mu_2}{\sigma_{\tau_2}^2} + \frac{1}{3\sigma^2} \sum_{\substack{\{(i,j,k) \mid \\ S_{i,j,k}^{(1,2,3)}=1\}}} (x_{1i} + A_1x_{2j} + \tau_1 - 2A_2x_{3k})}{\frac{1}{\sigma_{\tau_2}^2} + \frac{2m}{3\sigma^2}}, \frac{1}{\frac{1}{\sigma_{\tau_2}^2} + \frac{2m}{3\sigma^2}} \mathbf{I}_d \right). \end{aligned}$$

Distribuciones a priori de las matrices de rotación A_1 y A_2 y sus distribuciones a posteriori

Asumimos de nuevo, limitándonos al caso bidimensional, que las distribuciones a priori de las matrices de rotación A_1 y A_2 son distribuciones von Mises de parámetros ν_i y $k_i > 0$ con $i = 1, 2$, $A_i \sim M(\nu_i, k_i)$. Entonces, basándonos en el modelo (3.50), y procediendo como en el caso de transformaciones de similitud, la distribución a posteriori de A_1 es una distribución von Mises con función de densidad

$$p(A_1 \mid \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \tau_1, \tau_2, \sigma^2, S^{(1,2,3)}) \propto \exp \left\{ \text{traza} \left(F_1^{*T} A_1 \right) \right\}, \quad (3.52)$$

donde la matriz F_1^* toma la expresión

$$F_1^* = F_1 + \frac{1}{3\sigma^2} \sum_{\substack{\{(i,j,k) \mid \\ S_{i,j,k}^{(1,2,3)}=1\}}} (x_{1i} - 2\tau_1 + A_2x_{3k} + \tau_2)x_{2j}^T.$$

Análogamente, la distribución a posteriori de A_2 es una distribución von Mises con función de densidad

$$p(A_2 \mid \dots) \propto \exp \left\{ \text{traza} \left(F_2^{*T} A_2 \right) \right\}, \quad (3.53)$$

donde la matriz F_2^* es

$$F_2^* = F_2 + \frac{1}{3\sigma^2} \sum_{\substack{\{(i,j,k)| \\ S_{i,j,k}^{(1,2,3)}=1\}}} (x_{1i} + A_1 x_{2j} + \tau_1 - 2\tau_2) x_{3k}^T.$$

Para identificar los parámetros de estas distribuciones, llamamos

$$P^{(1)} = \begin{pmatrix} P_{11}^{(1)} & P_{12}^{(1)} \\ S_{21}^{(1)} & S_{22}^{(1)} \end{pmatrix} = \frac{1}{3\sigma^2} \sum_{\substack{\{(i,j,k)| \\ S_{i,j,k}^{(1,2,3)}=1\}}} (x_{1i} - 2\tau_1 + A_2 x_{3k} + \tau_2) x_{2j}^T,$$

y

$$P^{(2)} = \begin{pmatrix} P_{11}^{(2)} & P_{12}^{(2)} \\ P_{21}^{(2)} & P_{22}^{(2)} \end{pmatrix} = \frac{1}{3\sigma^2} \sum_{\substack{\{(i,j,k)| \\ S_{i,j,k}^{(1,2,3)}=1\}}} (x_{1i} + A_1 x_{2j} + \tau_1 - 2\tau_2) x_{3k}^T,$$

por lo que $F_i^* = F_i + P^{(i)}$ $i = 1, 2$.

Entonces, las expresiones (3.52) y (3.53) pueden escribirse

$$p(A_i | \dots) \propto \exp \{ \text{traza} ((F_i + P^{(i)})^T A_i) \}.$$

Así, se llega a que la distribución a posteriori de A_i , $i = 1, 2$ es von Mises de parámetros $A_i \sim M(\nu_i^*, k_i^*)$ donde

$$k_i^* = \left[\left(k_i \cos \nu_i + P_{11}^{(i)} + P_{22}^{(i)} \right)^2 + \left(k_i \sin \nu_i + P_{21}^{(i)} - P_{12}^{(i)} \right)^2 \right]^{1/2}$$

$$\nu_i^* = \arccos \left[\frac{k_i \cos \nu_i + P_{11}^{(i)} + P_{22}^{(i)}}{k_i^*} \right].$$

Distribución a priori de la matriz de emparejamientos $S^{(1,2,3)}$ y su estimación a posteriori mediante el algoritmo Metropolis-Hasting

A partir de (3.19) es inmediato que la distribución a priori de $S^{(1,2,3)}$, conociendo que hay m emparejamientos triples entre las tres configuraciones, es

$$p(S^{(1,2,3)} | M_{123}^{(3)} = m) = \frac{1}{(m!)^2},$$

es decir, la distribución uniforme sobre el total de posibles matrices de emparejamientos triples.

Se puede obtener una muestra de la distribución a posteriori utilizando el algoritmo de Metropolis-Hasting. A partir de este momento, y por simplicidad, incluimos una nueva notación para la expresión de la matriz de emparejamientos $S^{(1,2,3)}$.

Teníamos previamente que

$$S_{ijk}^{(1,2,3)} = \begin{cases} 1 & \text{si } \xi_{1i} = \xi_{2j} = \xi_{3k} \\ 0 & \text{resto} \end{cases}.$$

es decir, el array de dimensión tres $S^{(1,2,3)}$ toma el valor uno en aquellas ternas (ijk) que identifican los puntos x_{1i} , x_{2j} y x_{3k} que están emparejadas. Como partimos de que existen m emparejamientos, vamos a representarlos reuniéndolos en una matriz S de dimensiones $m \times 3$, cuyas filas recogerán aquellas ternas de puntos de las configuraciones $\mathbf{x}_1, \mathbf{x}_2$ y \mathbf{x}_3 que forman emparejamientos triples.

Así, por ejemplo, en el caso de $m = 5$, la matriz

$$S = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 3 \\ 3 & 4 & 2 \\ 4 & 5 & 5 \\ 5 & 3 & 4 \end{pmatrix}$$

está representando que los puntos de las tres configuraciones están emparejados de la siguiente manera: punto *uno* de la configuración \mathbf{x}_1 con el punto *uno* de la configuración \mathbf{x}_2 y con el punto *uno* de la configuración \mathbf{x}_3 , puntos *dos* de las configuraciones \mathbf{x}_1 y \mathbf{x}_2 con el tercer punto de la configuración \mathbf{x}_3 , el punto *tres* de la configuración \mathbf{x}_1 con el punto *cuatro* de la configuración \mathbf{x}_2 y con el punto *dos* de la configuración \mathbf{x}_3 , el punto *cuatro* de la configuración \mathbf{x}_1 con el punto *cinco* de la configuración \mathbf{x}_2 y con el punto *cinco* de la configuración \mathbf{x}_3 y, por último, el punto *cinco* de la configuración \mathbf{x}_1 con el punto *tres* de la configuración \mathbf{x}_2 y con el punto *cuatro* de la configuración \mathbf{x}_3 .

Es decir, está representando la situación

$$S_{111}^{(1,2,3)} = S_{223}^{(1,2,3)} = S_{342}^{(1,2,3)} = S_{455}^{(1,2,3)} = S_{534}^{(1,2,3)} = 1.$$

Se puede señalar que, en general, cada columna va a corresponder a una permutación de los números $1, 2, \dots, m$ y que el orden en el que aparezcan las filas, no es relevante.

Desarrollo del método basado en el algoritmo Metropolis-Hasting.

Las posibles transiciones permitidas en cada etapa de la cadena de Markov se han definido como sigue.

Se eligen dos puntos al azar de la configuración \mathbf{x}_1 , por ejemplo i_1 e i_2 , cuyos emparejamientos con los puntos de las configuraciones \mathbf{x}_2 y \mathbf{x}_3 los denotamos (i_1, j_1, k_1) y (i_2, j_2, k_2) que vendrán recogidos en la matriz S . Entonces, los posibles cambios permitidos de S a S^* con sus probabilidades respectivas $q(S, S^*)$ son:

- (i) Con probabilidad p_1^* intercambiar los emparejamientos sólo con la configuración \mathbf{x}_2 , de manera que el cambio en la matriz S consiste en sustituir los emparejamientos (i_1, j_1, k_1) e (i_2, j_2, k_2) por (i_1, j_2, k_1) e (i_2, j_1, k_2) , dando lugar a la matriz S^* .
- (ii) Con probabilidad p_2^* intercambiar los emparejamientos sólo con la configuración \mathbf{x}_3 , de manera que el cambio en la matriz S consiste en sustituir los emparejamientos (i_1, j_1, k_1) e (i_2, j_2, k_2) por (i_1, j_1, k_2) e (i_2, j_2, k_1) , dando lugar a una matriz S^* .
- (iii) Con probabilidad $1 - p_1^* - p_2^*$ intercambiar los emparejamientos con las configuraciones \mathbf{x}_2 y \mathbf{x}_3 , de manera que el cambio en la matriz S consiste en sustituir los emparejamientos (i_1, j_1, k_1) e (i_2, j_2, k_2) por (i_1, j_2, k_2) e (i_2, j_1, k_1) , respectivamente, dando lugar a una matriz S^* .

Estas probabilidades de transición de la cadena están definiendo la distribución propuesta o distribución de salto que representa el movimiento de la cadena en el

espacio de matrices de dimensión $m \times 3$. Hay que señalar que, tal como se ha definido esta distribución propuesta, la cadena verifica la propiedad de reversibilidad, es decir, la probabilidad de que partiendo de S en una etapa k , en la etapa $k + 1$ nos encontremos en S^* , coincide con la probabilidad de que partiendo de S^* en la etapa k en la etapa $k + 1$ nos encontremos en S , es decir,

$$q(S, S^*) = q(S^*, S).$$

En esta situación, y siguiendo con el desarrollo del método de Metropolis-Hasting, la probabilidad de aceptación del salto propuesto a la matriz S^* es,

$$\min \left\{ 1, r = \frac{p(A_1, A_2, \tau_1, \tau_2, \sigma^2, S^*, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)}{p(A_1, A_2, \tau_1, \tau_2, \sigma^2, S, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)} \right\}.$$

Observamos que r es el cociente entre el modelo conjunto, definido en (3.51), con los emparejamientos formados con la distribución propuesta y el modelo conjunto con los emparejamientos del paso anterior. Así la expresión de r será:

(i) Con probabilidad p_1^*

$$r = \frac{\exp \left\{ -\frac{1}{\sigma^2} \left(\frac{1}{4} \|x_{1i_1} - A_1 x_{2j_2} - \tau_1\|^2 + \frac{1}{3} \left\| \frac{x_{1i_1} + A_1 x_{2j_2} + \tau_1}{2} - A_2 x_{3k_1} - \tau_2 \right\|^2 \right. \right.}{\exp \left\{ -\frac{1}{\sigma^2} \left(\frac{1}{4} \|x_{1i_1} - A_1 x_{2j_1} - \tau_1\|^2 + \frac{1}{3} \left\| \frac{x_{1i_1} + A_1 x_{2j_1} + \tau_1}{2} - A_2 x_{3k_1} - \tau_2 \right\|^2 \right. \right.} \\ \left. \left. + \frac{1}{4} \|x_{1i_2} - A_1 x_{2j_1} - \tau_1\|^2 + \frac{1}{3} \left\| \frac{x_{1i_2} + A_1 x_{2j_1} + \tau_1}{2} - A_2 x_{3k_2} - \tau_2 \right\|^2 \right) \right\}}{\exp \left\{ -\frac{1}{\sigma^2} \left(\frac{1}{4} \|x_{1i_2} - A_1 x_{2j_2} - \tau_1\|^2 + \frac{1}{3} \left\| \frac{x_{1i_2} + A_1 x_{2j_2} + \tau_1}{2} - A_2 x_{3k_2} - \tau_2 \right\|^2 \right) \right\}}.$$

(ii) Con probabilidad p_2^*

$$r = \frac{\exp \left\{ -\frac{1}{\sigma^2} \left(\frac{1}{4} \|x_{1i_1} - A_1 x_{2j_1} - \tau_1\|^2 + \frac{1}{3} \left\| \frac{x_{1i_1} + A_1 x_{2j_1} + \tau_1}{2} - A_2 x_{3k_2} - \tau_2 \right\|^2 \right. \right.}{\exp \left\{ -\frac{1}{\sigma^2} \left(\frac{1}{4} \|x_{1i_1} - A_1 x_{2j_1} - \tau_1\|^2 + \frac{1}{3} \left\| \frac{x_{1i_1} + A_1 x_{2j_1} + \tau_1}{2} - A_2 x_{3k_1} - \tau_2 \right\|^2 \right. \right.} \\ \left. \left. + \frac{1}{4} \|x_{1i_2} - A_1 x_{2j_2} - \tau_1\|^2 + \frac{1}{3} \left\| \frac{x_{1i_2} + A_1 x_{2j_2} + \tau_1}{2} - A_2 x_{3k_1} - \tau_2 \right\|^2 \right) \right\}}{\exp \left\{ -\frac{1}{\sigma^2} \left(\frac{1}{4} \|x_{1i_2} - A_1 x_{2j_2} - \tau_1\|^2 + \frac{1}{3} \left\| \frac{x_{1i_2} + A_1 x_{2j_2} + \tau_1}{2} - A_2 x_{3k_2} - \tau_2 \right\|^2 \right) \right\}}.$$

(iii) Con probabilidad $1 - p_1^* - p_2^*$

$$r = \frac{\exp \left\{ -\frac{1}{\sigma^2} \left(\frac{1}{4} \|x_{1i_1} - A_1 x_{2j_2} - \tau_1\|^2 + \frac{1}{3} \left\| \frac{x_{1i_1} + A_1 x_{2j_2} + \tau_1}{2} - A_2 x_{3k_2} - \tau_2 \right\|^2 \right. \right.}{\exp \left\{ -\frac{1}{\sigma^2} \left(\frac{1}{4} \|x_{1i_1} - A_1 x_{2j_1} - \tau_1\|^2 + \frac{1}{3} \left\| \frac{x_{1i_1} + A_1 x_{2j_1} + \tau_1}{2} - A_2 x_{3k_1} - \tau_2 \right\|^2 \right. \right.} \\ \left. \left. + \frac{1}{4} \|x_{1i_2} - A_1 x_{2j_1} - \tau_1\|^2 + \frac{1}{3} \left\| \frac{x_{1i_2} + A_1 x_{2j_1} + \tau_1}{2} - A_2 x_{3k_1} - \tau_2 \right\|^2 \right) \right\}}{\left. + \frac{1}{4} \|x_{1i_2} - A_1 x_{2j_2} - \tau_1\|^2 + \frac{1}{3} \left\| \frac{x_{1i_2} + A_1 x_{2j_2} + \tau_1}{2} - A_2 x_{3k_2} - \tau_2 \right\|^2 \right) \right\}}.$$

En resumen, el algoritmo de Metropolis-Hasting para la estimación de la distribución a posteriori de la matriz de emparejamientos triples S , tendrá en el paso de la iteración k a la $k + 1$, las siguientes acciones:

- a. Elegir aleatoriamente dos números del 1 al m , que representarán los puntos de la configuración \mathbf{x}_1 que van a intercambiar emparejamientos.
- b. Elegir un valor para p_1^* y para p_2^* (tomaremos p_1^* y $p_2^* = 1/3$) con los que proponer los emparejamientos en la etapa siguiente.
- c. Calcular la probabilidad de aceptación del salto propuesto r , definida como el cociente entre el modelo conjunto con los emparejamientos definidos con la matriz S en el paso $k + 1$, y el modelo conjunto con los emparejamientos en el paso k . Si r es mayor que 1, el salto propuesto se acepta. Si es menor, se acepta con probabilidad r .

3.7. Aplicaciones

En esta sección se presentan dos tipos de aplicaciones, que ilustran los resultados presentados en este capítulo. Por una parte se validan los resultados teóricos expuestos en las secciones anteriores a través de configuraciones simuladas y por otro se realiza una aplicación en el ámbito de la Bioinformática, utilizando la base

de datos de microarrays de Karaman et al (2003) basada en la plataforma *Affymetrix Genechip 5.0*. Para todo ello se ha utilizado principalmente MatLab para programar las simulaciones y la plataforma Bioconductor de R para obtener los datos reales, con un ordenador Pentium IV con procesador a 3.2 Ghz.

3.7.1. Datos simulados

Como primer paso a la comprobación de la validez de los resultados teóricos, se han simulado datos de un proceso de Poisson espacial en un círculo en \mathbb{R}^2 (tomando un radio igual a 8) y de tasa $\lambda = 2$ siguiendo el algoritmo de Ross (1997) presentado en la sección 2.2. A partir de él, y siguiendo el modelo (3.1), se han generado por una parte $M = 4$ configuraciones con transformaciones rígidas, y por otra $M = 3$ configuraciones con transformaciones de similitud, en ambas situaciones suponiendo conocidos los emparejamientos. Los valores de los vectores de traslación, de las matrices de rotación y de la varianza de los errores utilizados para estas simulaciones se muestran en la tablas 3.1. y 3.2.

En el caso primer caso, se ha aplicado un muestreador de Gibbs con 50000 iteraciones en total, 10000 de ellas de calentamiento, para generar muestras aleatorias de la distribuciones a posteriori de los parámetros $A_1, A_2, A_3, \tau_1, \tau_2, \tau_3$ y σ^2 . Se han tomado como estimaciones de A_1, A_2 y A_3 , las matrices de rotación de los ángulos medios de las distribuciones a posteriori y para el resto de los parámetros las medias a posteriori. La programación de los codes se ha hecho utilizando Matlab v. 7.1. El tiempo aproximado de simulación en este primer caso ha sido de 2430 segundos (40.5 minutos).

En el caso de $M = 3$ configuraciones con transformaciones de similitud, la generación de las muestras de la distribución a posteriori de las matrices de rotación, vectores de traslación y varianza se obtuvo mediante un muestreador de Gibbs. Las muestras a posteriori de los parámetros de escala s_1 y s_2 se generaron mediante el *método strip*, que consiste básicamente en generar una muestra de la distribución

discreta definida sobre una rejilla de puntos en el dominio de la distribución a generar; a cada punto se le asigna como probabilidad el cociente de la función de densidad en ese punto y la suma de los valores de las funciones de densidades en todos los puntos de la rejilla. (Para más detalles ver Devroye, 1986). El tiempo aproximado de simulación en este segundo caso ha sido de 19404 segundos (5.39 horas).

En ambos casos, se tomaron como distribuciones a priori

$$\begin{aligned}\tau_i &\sim \mathbf{N}_2((2, 2)^T, \mathbf{I}_2) \\ \sigma^{-2} &\sim G(1, 10) \\ A_i &\sim M(u, 2) \text{ von Mises } (u \text{ un valor aleatorio de } [0, 2\pi]) \\ s_i &\sim G(1, 0,002)\end{aligned}$$

En las tablas 3.1 y 3.2 se comparan las estimaciones obtenidas en la simulación y los valores reales de los parámetros utilizados en la generación de las configuraciones en ambas situaciones. Se puede comprobar que, salvo algunos valores de los parámetros de traslación donde no se ajustan tan bien las estimaciones, la aproximación de las simulaciones a los valores reales es muy exacta.

Valores de los parámetros	Estimaciones simuladas
$\tau_1 = [2, 3]$ $\tau_2 = [1, 1]$ $\tau_3 = [0, 0]$	$\hat{\tau}_1 = [1,8337, 3,2990]$ $\hat{\tau}_2 = [1,1295, 0,6218]$ $\hat{\tau}_3 = [0,0670, 0,0138]$
$\theta_1 = 30^\circ \quad A_1 = \begin{pmatrix} 0,866 & -0,500 \\ 0,5 & 0,866 \end{pmatrix}$	$\hat{\theta}_1 = 31,3^\circ \quad \hat{A}_1 = \begin{pmatrix} 0,8535 & -0,521 \\ 0,521 & 0,8535 \end{pmatrix}$
$\theta_2 = 60^\circ \quad A_2 = \begin{pmatrix} 0,5 & -0,866 \\ 0,866 & 0,5 \end{pmatrix}$	$\hat{\theta}_2 = 59,3^\circ \quad \hat{A}_2 = \begin{pmatrix} 0,5093 & -0,8606 \\ 0,8606 & 0,5093 \end{pmatrix}$
$\theta_3 = 45^\circ \quad A_3 = \begin{pmatrix} 0,7071 & -0,7071 \\ 0,7071 & 0,7071 \end{pmatrix}$	$\hat{\theta}_3 = 45,8^\circ \quad \hat{A}_3 = \begin{pmatrix} 0,6964 & -0,7177 \\ 0,7177 & 0,6964 \end{pmatrix}$
$\sigma = 3$	$\hat{\sigma} = 2,943$

Tabla 3.1. Datos simulados en el caso de $M = 4$ configuraciones de puntos emparejados con transformaciones rígidas

Valores de los parámetros	Estimaciones simuladas
$\tau_1 = [2, 3]$ $\tau_2 = [2, 3]$	$\hat{\tau}_1 = [1,9949, 2,9475]$ $\hat{\tau}_2 = [1,7443, 3,1233]$
$s_1 = 2$ $s_2 = 2$	$\hat{s}_1 = 2,0303$ $\hat{s}_2 = 1,9900$
$\theta_1 = 30^\circ$ $A_1 = \begin{pmatrix} 0,8666 & -0,5 \\ 0,5 & 0,8666 \end{pmatrix}$	$\hat{\theta}_1 = 28,6^\circ$ $\hat{A}_1 = \begin{pmatrix} 0,8766 & -0,4812 \\ 0,4812 & 0,8766 \end{pmatrix}$
$\theta_2 = 60^\circ$ $A_2 = \begin{pmatrix} 0,5 & -0,8666 \\ 0,8666 & 0,5 \end{pmatrix}$	$\hat{\theta}_2 = 62,6^\circ$ $\hat{A}_2 = \begin{pmatrix} 0,4593 & -0,8883 \\ 0,8883 & 0,4593 \end{pmatrix}$
$\sigma = 2$	$\hat{\sigma} = 2,1883$

Tabla 3.2. Datos simulados en el caso de $M = 3$ configuraciones de puntos emparejados con transformación de similitud

3.7.2. Una aplicación en Bioinformática

La aplicación que aquí se presenta se basa en los datos tomados de un experimento con microarrays de Karaman et al. (2003) con células cultivadas de fibroblastos en tres especies genéticamente relacionadas: gorilas (*Gorilla gorilla*), bonobos (*Pan paniscus*) y humanos (*Homo sapiens*). Dichos datos se pueden obtener a partir del paquete *fibroEset* de Bioconductor.

Estos datos representan los valores de las expresiones de un total de 12925 genes en 46 muestras (23 de humanos, 11 de bonobos y 12 de gorilas). Se asume que aquellos genes que se muestran más expresados serán aquellos que tienen mayor interés, de modo que se han seleccionado aquellos con una expresión mayor de 3000, lo que supone una selección de 204 genes.

En el cuadro 3.1 se representa el aspecto de la base de datos con la que se va a

trabajar, donde \exp_{ij}^k representa la expresión del gen i en la muestra j de la especie k ($k \in \{\text{Humano, Bonobo, Gorila}\}$). Las filas representan los genes seleccionados y las columnas las 46 muestras identificando la especie de la que proceden (H: humano, G: gorila y B:bonobo).

	H1	...	H23	G1	...	G11	B1	...	B12
Gen 1	$\exp_{1,1}^H$...	$\exp_{1,23}^H$	\exp_{11}^B	...	$\exp_{1,11}^B$	\exp_{11}^G	...	$\exp_{1,12}^G$
\vdots	\vdots		\vdots	\vdots		\vdots	\vdots		\vdots
Gen 204	$\exp_{204,1}^H$...	$\exp_{204,23}^H$	$\exp_{204,1}^B$...	$\exp_{204,11}^B$	$\exp_{204,1}^G$...	$\exp_{204,12}^G$

Cuadro 3.1 Esquema de la base de datos de microarrays

Calculando las distancias euclídeas entre los genes y aplicando la técnica multi-variante INDSCAL, se proyectan los genes de cada una de las tres especies en un espacio de dimension $r = 2$. Estas representaciones son las que jugarán el papel de configuraciones de puntos. Al conocerse qué punto representa cada gen y al haber seleccionado los mismos genes en las tres especies, se tiene tres configuraciones de 204 puntos cuyos emparejamientos son conocidos, por lo que se planteará la estimación del modelo bajo el supuesto de configuraciones etiquetadas.

Por último, se trabajará con la hipótesis de que no se conocen los emparejamientos entre genes (configuraciones no etiquetadas), por lo que se deberán realizar estimaciones de los emparejamientos. En ambos casos se supondrán transformaciones rígidas entre configuraciones, por lo que estos problemas son aplicaciones de los apartados 3.6.3 y 3.6.4 del presente capítulo.

Representación de los genes.

La representación de los genes se ha llevado a cabo utilizando un análisis INDSCAL (INDividual Differences SCALing) o escalamiento multidimensional de diferencias individuales (Carroll y Chang, 1970), presentado en el capítulo 2.

En nuestro caso, los genes hacen el papel de objetos y las especies hacen papel de individuos. Así, el análisis INDSCAL permite tener en cuenta las diferencias entre especies a la hora de representar los genes, de tal modo que se obtendrá una representación de los genes distinta para cada especie según las ponderaciones que ofrezca el análisis.

En primer lugar, se han tomado como matrices de similitud las matrices de las distancias euclídeas entre los genes para cada especie calculadas a partir de sus expresiones, que servirán de entrada al análisis INDSCAL.

El resultado del análisis proporciona, por una parte, las coordenadas de todos los genes sin diferenciar por especie en el espacio global de objetos o de estímulos, (de dimensión $r = 2$), y por otra, los pesos o ponderaciones que cada especie asigna a cada dimensión, representados en el espacio de sujetos.

La lectura y tratamiento de las expresiones de los genes han sido realizados por el programa Bioconductor y el análisis INDSCAL de los datos se ha llevado a cabo con el procedimiento MDS de SAS v 9.1.

En las figuras 3.1 y 3.2 se representan los dos espacios obtenidos con los 204 genes. Se han identificado dos genes con su etiqueta únicamente a efectos ilustrativos y comparativos. En el espacio de sujetos se observa que los gorilas dan más importancia a la dimensión 2 que los humanos y bonobos, y que las ponderaciones que aplican estos últimos son más similares que las que aplican los gorilas. Esto se refleja en las figuras 3.3, 3.4 y 3.5 donde aparecen representados las tres configuraciones de genes.

CAPÍTULO 3. EMPAREJAMIENTO DE MÁS DE DOS CONFIGURACIONES NO ETIQUETADAS MEDIANTE TRANSFORMACIONES LINEALES

Se observa que la representación más dispar es la correspondiente a los gorilas.

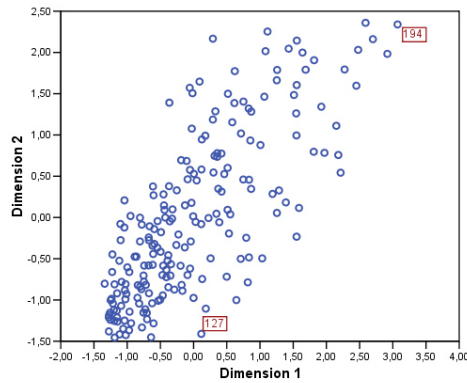


Figura 3.1. Espacio global de objetos

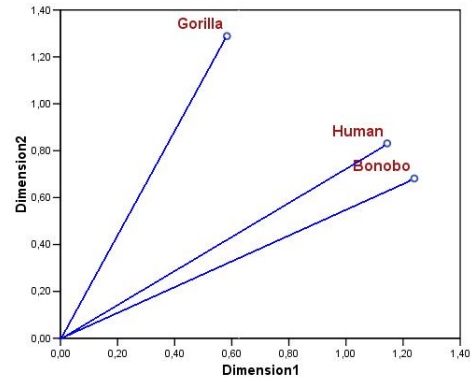


Figura 3.2. Espacio de sujetos

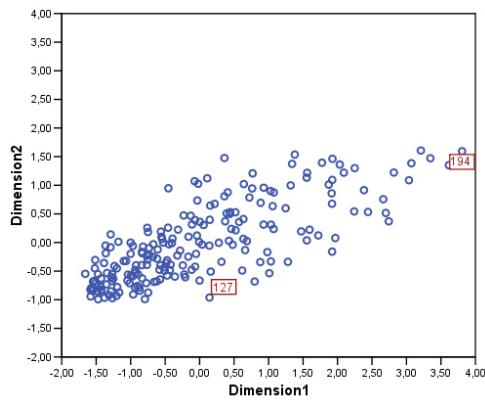


Figura 3.3 Configuración de bonobos

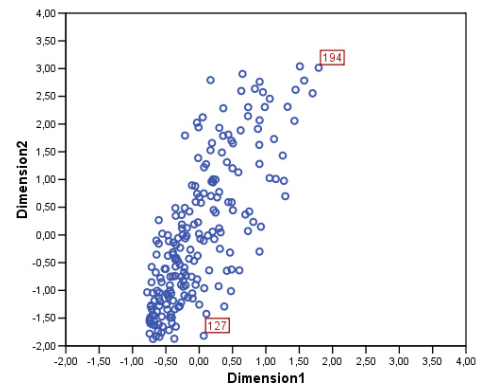


Figura 3.4. Configuración de gorilas

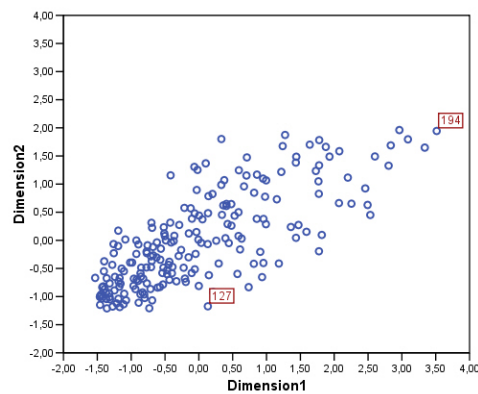


Figura 3.5. Configuración de humanos

Estimación de la transformación rígida entre especies con emparejamientos conocidos de los genes.

Al haber seleccionado los mismos genes en las tres especies y conocer a qué gen corresponde cada punto, se tienen 204 emparejamientos triples y ninguno doble y además se conoce cuáles son los puntos que componen los emparejamientos, por lo que las matrices de afinidad y de emparejamiento son conocidas. Por lo tanto el problema se centra en encontrar las transformaciones rígidas (matrices de rotación y vectores de traslación) que relacionan las configuraciones, además de la varianza de los errores.

Análogamente al caso de datos simulados y de nuevo utilizando Matlab 7.1, se han generado muestras aleatorias de las distribuciones a posteriori de los parámetros A_1 , A_2 , τ_1 , τ_2 y σ^2 aplicando un muestreador de Gibbs con 50000 iteraciones en total, 10000 de ellas de calentamiento. Inicialmente se considera como configuración de referencia la de los humanos obteniéndose las transformaciones entre humanos y bonobos y gorilas; en un segundo paso se ha tomado la de los gorilas como referencia para también conseguir las transformaciones entre gorilas y bonobos.

Así la rotación y traslación entre las configuraciones de humanos y bonobos viene dada por:

$$A_1 = \begin{pmatrix} 0,9949 & -0,1008 \\ 0,1008 & 0,9949 \end{pmatrix}, \quad \tau_1 = \begin{pmatrix} 0,0015 \\ 0,0013 \end{pmatrix}$$

lo que indica una rotación de la configuración de bonobos a humanos de $5,78^\circ$ (0,10 radianes) y ninguna traslación.

La transformación entre las configuraciones de humanos y gorilas es:

$$A_2 = \begin{pmatrix} 0,9148 & 0,4039 \\ -0,4039 & 0,9148 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0,0014 \\ 0,0012 \end{pmatrix}$$

lo que significa una rotación de $-23,82^\circ$ (5,86 radianes) de la configuración de los gorilas a los humanos. De nuevo no es necesaria ninguna traslación.

La rotación de gorilas a bonobos es de $29,61^\circ$ (0,51 radianes) y ninguna traslación.

Las estimaciones de la matriz de rotación y del vector de traslación son las siguientes:

$$A_3 = \begin{pmatrix} 0,8694 & -0,4941 \\ 0,4941 & 0,8694 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 0,0013 \\ 0,0014 \end{pmatrix}.$$

Por último, la estimación de la varianza obtenida es $\sigma^2 = 0,0446$.

Descripción de las distribuciones a posteriori y análisis de las cadenas.

En la tabla 3.3 se muestran algunos estadísticos básicos de las muestras generadas de las distribuciones a posteriori eliminada su fase de *burning*. En particular, se han calculado la media, mediana y desviación típica. También se muestran los intervalos de mayor probabilidad al 95 % y se representan las densidades de las distribuciones a posteriori de todos los parámetros (figuras 3.6 hasta figura 3.15).

	Media	Mediana	Desviación típica	Ext. inf. int. 95 %	Ext. sup. int. 95 %
σ^2	0.04454	0.04449	0.00222	0.04025	0.04889
τ_{11}	0.00125	0.00133	0.02089	-0.03887	0.04259
τ_{12}	0.00138	0.00145	0.02106	-0.04079	0.04207
τ_{21}	0.00116	0.00132	0.02085	-0.04011	0.04137
τ_{22}	0.00141	0.00140	0.02085	-0.04072	0.04114
τ_{31}	0.00131	0.00119	0.02085	-0.03858	0.04314
τ_{32}	0.00137	0.00140	0.02089	-0.03892	0.042906
θ_1	5.86734	5.86730	0.01508	5.8376	5.8966
θ_2	0.10115	0.10111	0.01489	0.0713	0.1295
θ_3	0.5164	0.51644	0.01519	0.4866	0.5459

Tabla 3.3 Análisis descriptivo de las cadenas

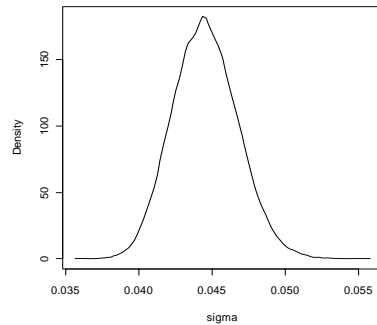


Fig. 3.6 Densidad de sigma

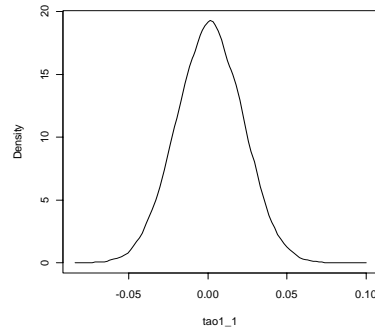


Fig. 3.7 Densidad 1ª coordenada de τ_1

CAPÍTULO 3. EMPAREJAMIENTO DE MÁS DE DOS CONFIGURACIONES NO ETIQUETADAS MEDIANTE TRANSFORMACIONES LINEALES

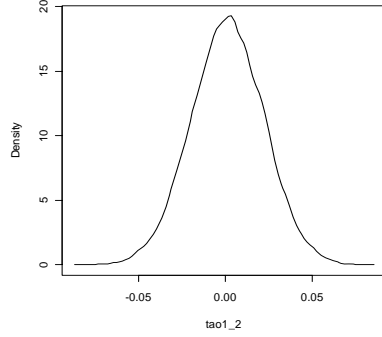


Fig. 3.8 Densidad 2ª coordenada de τ_1

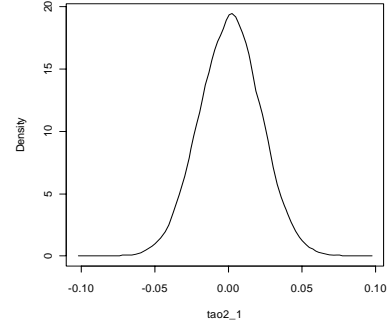


Fig. 3.9 Densidad 1ª coordenada de τ_2

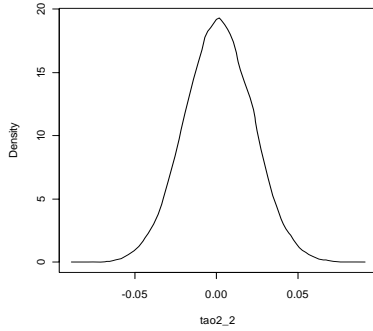


Fig. 3.10 Densidad 2ª coordenada de τ_2

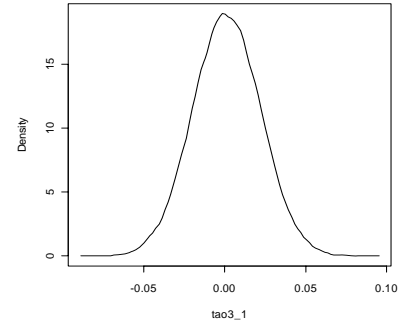


Fig. 3.11 Densidad 1ª coordenada de τ_3

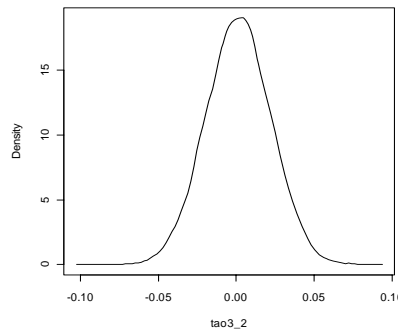


Fig. 3.12 Densidad 2ª coordenada de τ_3

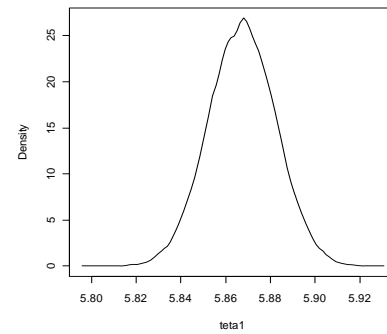


Fig. 3.13 Densidad de θ_1

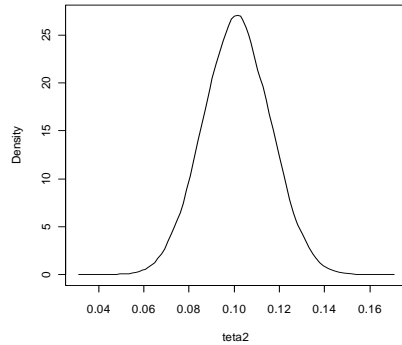


Fig. 3.14 Densidad de θ_2

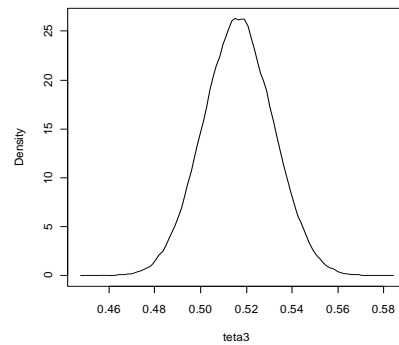


Fig. 3.15 Densidad de θ_3

Para estudiar la convergencia de las cadenas se han utilizado los tests de Geweke y el de Raftery y Lewis. El primero de ellos contrasta la estacionariedad de la cadena y el segundo la independencia de los valores de la misma. Los dos son tests de diagnóstico aplicados a una sola cadena y se han calculado con el paquete *boa* de R.

El test de Geweke (1992) considera dos subconjuntos de una cadena (habitualmente el primer 10 % y el último 50 %, obviando la parte de *burning*), y se comparan las medias muestrales de los dos grupos. Cuando se obtiene un valor no significativo, se acepta la igualdad entre ambas medias y, por lo tanto, no se tiene evidencia de que falle la convergencia.

El test de Raftery y Lewis (1992), tiene como objetivo calcular la longitud del periodo de *burning* y el número total de iteraciones necesarias para estimar un cuantil dado de la distribución a posteriori. Frecuentemente se toma el cuantil 0.025 y se debe especificar la precisión requerida para su estimación y una probabilidad prefijada de que el cuantil estimado tenga la precisión dada. Se define el factor de dependencia como el número total de iteraciones dividido entre el número mínimo de iteraciones necesarias para la estimación del cuantil. Un valor mayor que 5 de este factor de dependencia, indica correlación entre los valores de la cadena.

En la tabla 3.4 se presentan los resultados obtenidos para cada uno de los parámetros del modelo. Los valores del estadístico de Geweke se resumen en la primera

CAPÍTULO 3. EMPAREJAMIENTO DE MÁS DE DOS CONFIGURACIONES NO ETIQUETADAS MEDIANTE TRANSFORMACIONES LINEALES

columna acompañados de los p -valores. Se observa que no se tiene evidencia suficiente de que falla la convergencia para ninguno de los parámetros. Los factores de dependencia del test de Raftery y Lewis son menores que cinco, lo que indica que no existe una correlación significativa. Por último, se muestran los gráficos de las trazas de todas las cadenas (figuras 3.16 hasta figura 3.25).

	Test Geweke		Test Raftery y Lewis
	Z-score	p-valor	Factor Dependencia
σ^2	-0.42427	0.6713	1.0045
τ_{11}	-1.1941	0.2324	1.9791
τ_{12}	0.4394	0.6603	2.1601
τ_{21}	-0.1570	0.8752	1.9845
τ_{22}	0.4851	0.6275	1.9071
τ_{31}	0.7215	0.4706	2.0768
τ_{32}	0.3287	0.7423	2.1062
θ_1	-0.9607	0.3366	2.0411
θ_2	0.3175	0.7507	2.0469
θ_3	0.1989	0.8423	2.0405

Tabla 3.4 Tests de diagnóstico de convergencia

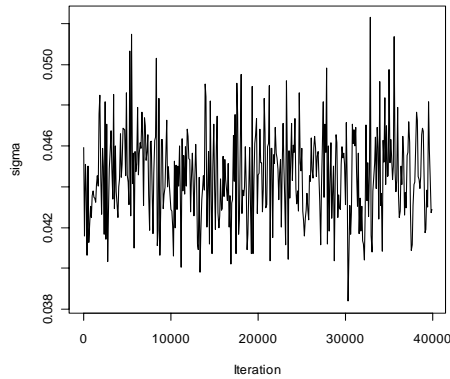


Fig. 3.16 Traza de sigma

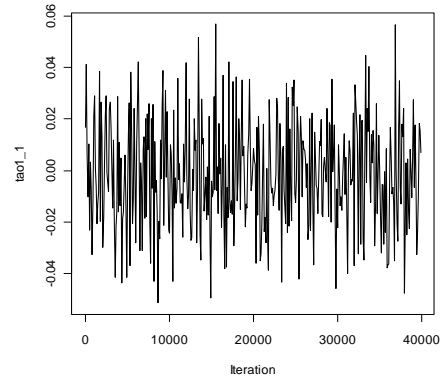


Fig. 3.17 Traza 1ª coordenada de τ_1

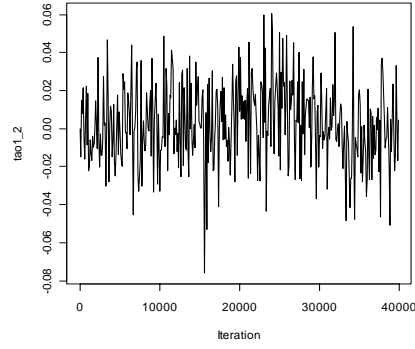


Fig. 3.18 Traza 2^a coordenada de τ_1

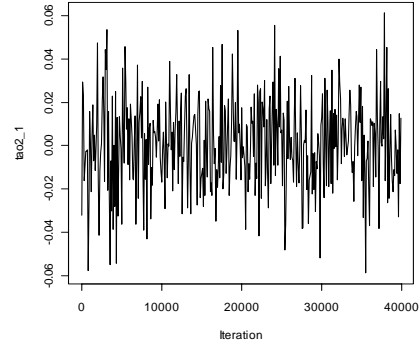


Fig. 3.19 Traza 1^a coordenada de τ_2

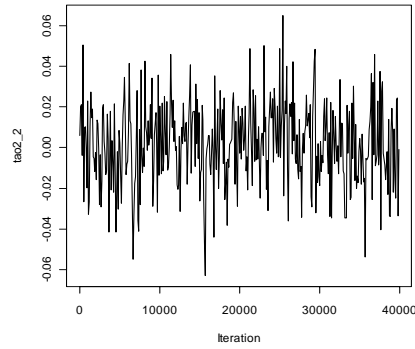


Fig. 3.20 Traza 2^a coordenada de τ_2

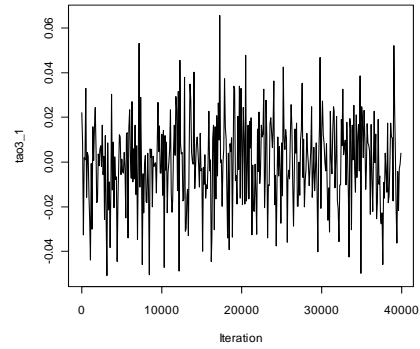


Fig. 3.21 Traza 1^a coordenada de τ_3

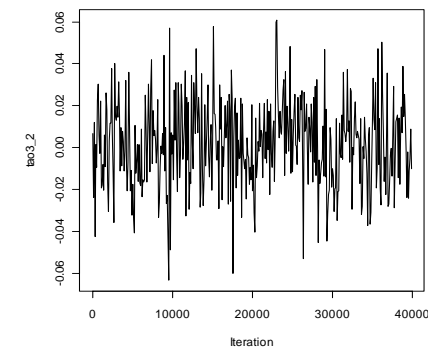


Fig. 3.22 Traza 2^a coordenada de τ_3

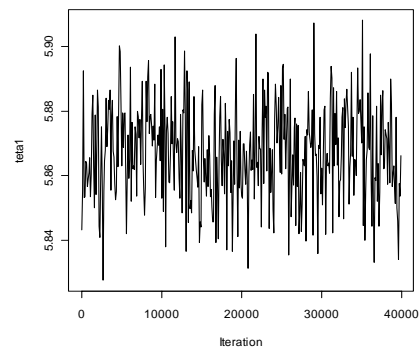


Fig. 3.23 Traza de θ_1

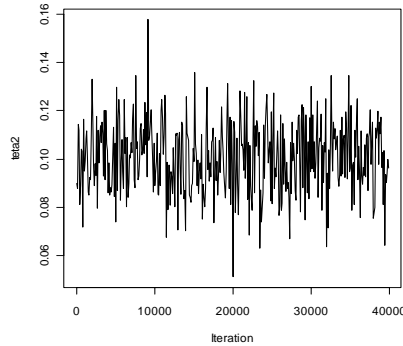


Fig. 3.24 Traza de θ_2

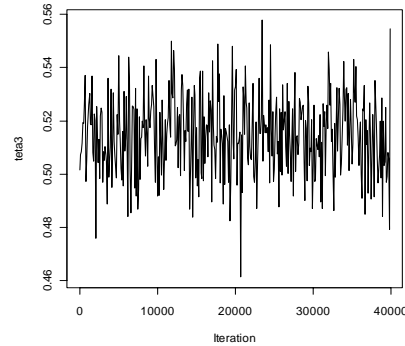


Fig. 3.25 Traza de θ_3

Como conclusión se observa que las tres especies están altamente relacionadas genéticamente, siendo mayor esta relación entre bonobos y humanos incluso que la existente entre gorilas y bonobos. Este resultado es consistente con la representación del árbol filogenético de los homínidos. Se sabe que los datos de comparación de muestras de ADN muestran una similitud del 98.5% entre el ADN humano y el chimpacé y esta semejanza es mayor que la que existe entre chimpacé y gorila o gorila y humano.

Como futuras aplicaciones, se puede aplicar esta metodología sobre otros genes relevantes desde el punto de vista metabólico y encontrar relaciones existentes entre diferentes organismos. Además se pueden plantear otros tipos de distancias, ya que el análisis INDSCAL en particular, y el multidimensional scaling en general, se basa en la matriz de similitud entre puntos, pudiéndose utilizar distintas distancias.

Estimación de los emparejamientos entre genes.

Como segunda aplicación del desarrollo presentado en este capítulo, se va a validar el procedimiento de estimación de las matrices de emparejamiento seleccionando ciertos genes de la base de datos. Para ello, aunque realmente se conocen los emparejamientos, se va a suponer que son desconocidos de manera que utilizaremos el método Metropolis-Hasting desarrollado en la sección 3.6.4 para la estimación de dichos emparejamientos.

Se considera la misma base de datos y se seleccionan aquellos genes con expresiones entre 8000 y 10000, obteniendo un subconjunto de 23 genes. De nuevo se representan en dos dimensiones utilizando la técnica INDSCAL y se supone que se desconocen las identificaciones de los puntos, de manera que no se sabe qué gen representa cada punto y, por lo tanto, necesitamos emparejarlos para, al menos, descubrir qué puntos de las tres configuraciones identifican a los mismos genes.

Para ello se emplea un método empírico Bayes (Carlin y Louis, 2000) que consiste, básicamente, en realizar una estimación previa de las matrices de rotación con otro subconjunto de genes para los que sí se conocen los emparejamientos. Una vez estimadas, se utilizan en el proceso para estimar los emparejamientos triples del conjunto de genes de interés.

Seleccionamos una muestra de genes con expresiones mayores que 10000 (en total, 38 genes). Se aplica la técnica INDSCAL, y dado que los 38 puntos están emparejados y son conocidos, se estiman las matrices de rotación de humanos a bonobos y de humanos a gorilas, aplicando el MCMC con 60000 iteraciones en total, 20000 de ellas de calentamiento.

Los resultados obtenidos son,

$$A_1 = \begin{pmatrix} 0,9781 & 0,2081 \\ -0,2081 & 0,9781 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0,9906 & -0,1366 \\ 0,1366 & 0,9906 \end{pmatrix},$$

es decir, una rotación de -12.01° (-0.21 radianes) de humanos a bonobos y de 7.84° (0.1370 radianes) de humanos a gorilas.

A continuación se aplica el método Metropolis-Hasting descrito en 3.6.4 para la estimación de los emparejamientos de los 23 genes de interés, suponiendo conocidas las matrices de rotación e iguales a las estimaciones obtenidas. La matriz de emparejamientos $S^{(1,2,3)}$ se estimará seleccionando los 23 emparejamientos no repetidos más frecuentes de la muestra a posteriori obtenida. La programación de los codes se ha hecho utilizando Matlab v. 7.1.

En la base de datos original, cuyo esquema se mostraba en el cuadro 3.1, cada

fila recoge las expresiones del mismo gen en las tres especies, por lo que a la hora de calcular sus coordenadas en las configuraciones, el análisis INDSCAL también identifica los puntos con el mismo orden, es decir, en la base de datos donde se tienen las coordenadas de los genes de las tres configuraciones calculadas con el análisis INDSCAL, la fila k corresponde a las coordenadas del gen k en bonobos, gorilas y humanos. Esto implica que realmente, están emparejados los puntos x_{1k} , x_{2k} y x_{3k} , aunque estamos suponiendo que no se conocen. De esta manera, se comprobará la validez del método si el procedimiento identifica que los puntos k ($k = 1, \dots, 23$) de las tres configuraciones son los que están emparejados.

Los resultados obtenidos después de 60000 iteraciones, 20000 de calentamiento, se muestran en la tabla 3.5. En ella se observa cómo el procedimiento ha identificado correctamente los genes que están emparejados. Además se recogen las frecuencias relativas de cada emparejamiento en la muestra a posteriori simulada. El emparejamiento que más veces ha aparecido en esta muestra es el del gen 23 en las tres especies (con una frecuencia del 66 % de las veces). Las estimaciones obtenidas de los vectores de traslación fueron $(0.2312, 0.2341)^T$ y $(0.2309, 0.2356)^T$ y la estimación de la varianza fue 1.01.

	Número de los Genes emparejados									
Humanos	23	1	2	3	4	9	5	7	6	22
Bonobos	23	1	2	3	4	9	5	7	6	22
Gorilas	23	1	2	3	4	9	5	7	6	22
Frecuencia	0.66	0.49	0.44	0.43	0.43	0.42	0.41	0.40	0.39	0.39

Humanos	10	21	14	15	16	8	11	18	19	17
Bonobos	10	21	14	15	16	8	11	18	19	17
Gorilas	10	21	14	15	16	8	11	18	19	17
Frecuencia	0.39	0.37	0.37	0.37	0.37	0.39	0.38	0.38	0.38	0.38

Humanos	20	12	13
Bonobos	20	12	13
Gorilas	20	12	13
Frecuencia	0.38	0.38	0.38

Tabla 3.5. Lista de genes emparejados. Los genes número 23 de las tres especies forman el emparejamiento más frecuente (66 %)

CAPÍTULO 3. EMPAREJAMIENTO DE MÁS DE DOS CONFIGURACIONES NO ETIQUETADAS MEDIANTE TRANSFORMACIONES LINEALES

Los contenidos de este capítulo 3, incluida la aplicación en Bioinformática, puede encontrarse publicados en Marín y Nieto (2008a).

Capítulo 4

Emparejamiento de pares de configuraciones mediante transformaciones no lineales

4.1. Introducción

En muchas ocasiones, debido a la complejidad de las estructuras de las configuraciones, el problema de su emparejamiento puede ser de gran dificultad. En esta situación, la hipótesis de que la transformación existente que relaciona los emparejamientos es lineal puede suponerse demasiado restrictiva.

En este capítulo se va a presentar el análisis del emparejamiento de dos configuraciones a través de modelos no lineales. En particular, se elige como transformación no lineal el modelo de redes neuronales que se introdujo en el capítulo 2.

En una primera sección se desarrollará el modelo de Green y Mardia (2006) bajo estas hipótesis, donde las dos configuraciones son aleatorias y generadas a partir de un proceso de Poisson, con emparejamientos desconocidos, presentando los resultados teóricos que se obtienen en este contexto de no linealidad entre configuraciones no etiquetadas.

En una siguiente sección se plantea un segundo modelo en el que nos centramos en configuraciones etiquetadas y se considera fija a una de las dos configuraciones, siendo la otra, una transformación no lineal de la primera más un error aleatorio.

Para finalizar, se presenta una aplicación de ambos modelos en el ámbito de la Bioinformática, utilizando la base de datos ALL (Acute Lymphoblastic Leukemia) de Chiaretti et al. (2004), que recoge las características y expresiones de genes de 128 pacientes enfermos de leucemia.

4.2. Emparejamiento de configuraciones no etiquetadas bajo modelo de redes neuronales basado en un Proceso de Poisson

La situación de la que partimos es la descrita en la sección 3.2, pero limitada a dos configuraciones y con transformaciones no lineales entre los espacios donde se representan. Al centrarnos en dos configuraciones, el modelo que se desarrolla coincide parcialmente con el de Green y Mardia (2006), presentado en la sección 2.4, pero con la suposición de que la transformación geométrica es no lineal.

4.2.1. Modelización del problema

Sean las configuraciones \mathbf{x} e \mathbf{y} situadas en el espacio \mathbb{R}^d , donde $d \geq 2$, con n y m puntos cada una respectivamente, es decir,

$$\mathbf{x} = \{x_i, i = 1, \dots, n\} \quad \mathbf{y} = \{y_j, j = 1, \dots, m\}.$$

Las configuraciones son perturbaciones aleatorias de un conjunto de puntos fijos y desconocidos $\{\mu_l\} \in \mathbb{R}^d$ $l = 1, \dots, N$, la configuración de referencia, y cada μ_l sólo puede generar o un punto de \mathbf{x} , o un punto de \mathbf{y} , o un punto de cada una de ellas o ningún punto. Se desconoce qué puntos de cada configuración se corresponden con cada uno de estos μ_l , y se tiene que aquellos puntos de \mathbf{x} e \mathbf{y} que son perturbaciones

aleatorias de un mismo punto μ_l están emparejados. En la figura 4.1 se representa la situación entre configuraciones.

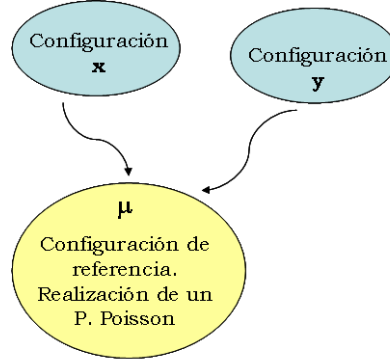


Figura 4.1 Esquema modelo con Proceso Poisson

La correspondencia entre los puntos μ_l y los puntos de las configuraciones, se representa mediante los vectores ξ_{1i} , $i = 1, \dots, n$ y ξ_{2j} , $j = 1, \dots, m$. Así, ξ_{1i} representa el subíndice de μ_l que genera el punto i de la configuración \mathbf{x} y, análogamente, ξ_{2j} representa el subíndice de μ_l que genera el punto j de la configuración \mathbf{y} .

Además, existe una transformación no lineal desconocida entre los sistemas de referencia donde se representan las dos configuraciones, cuyos parámetros también formarán parte del modelo. De nuevo, y sin pérdida de generalidad, consideramos la configuración \mathbf{x} en el sistema de referencia donde se encuentra $\{\mu_l\}$. Así, los puntos de \mathbf{x} se generan a partir de los $\{\mu_l\}$ más un error aleatorio.

Bajo estos supuestos, el modelo matemático es un caso particular de (3.1) y queda de la forma,

$$\begin{aligned} x_i &= \mu_{\xi_{1i}} + \varepsilon_i & i &= 1, \dots, n \\ \phi(y_j) &= \mu_{\xi_{2j}} + \epsilon_j & j &= 1, \dots, m, \end{aligned}$$

donde $\varepsilon_i \sim f_1 \in \mathbb{R}^d$ y $\epsilon_j \sim f_2 \in \mathbb{R}^d$, $i = 1, \dots, n$ $j = 1, \dots, m$ independientes, y

$$\phi(y_j) = \begin{pmatrix} \beta_{10} + \sum_{k=1}^M \beta_{1k} \Psi(\gamma_{k0} + y_j^T \gamma_k) \\ \vdots \\ \beta_{r0} + \sum_{k=1}^M \beta_{rk} \Psi(\gamma_{k0} + y_j^T \gamma_k) \\ \vdots \\ \beta_{d0} + \sum_{k=1}^M \beta_{dk} \Psi(\gamma_{k0} + y_j^T \gamma_k) \end{pmatrix},$$

siendo los parámetros del modelo $\beta_{j0} \in \mathbb{R}$, $\beta_{jk} \in \mathbb{R}$, $\gamma_{k0} \in \mathbb{R}$ y $\gamma_k \in \mathbb{R}^d$ para todo $j = 1, \dots, d$ y $k = 1, \dots, M$, y $\Psi(z) = \frac{\exp(z)}{1+\exp(z)}$ la función logística. El valor de M representa el número de nodos de la red neuronal y, en principio, lo consideraremos fijo y conocido, aunque en la aplicación con datos reales, se llevará a cabo un análisis de los modelos con distintos valores de M basado en criterios de información.

Entonces, el modelo se puede expresar como:

$$\begin{aligned} x_i &= \mu_{\xi_{1i}} + \varepsilon_i & i = 1, \dots, n & \quad \varepsilon_i \sim f_1 \in \mathbb{R}^d \\ \begin{pmatrix} \beta_{10} + \sum_{k=1}^M \beta_{1k} \Psi(\gamma_{k0} + y_j^T \gamma_k) \\ \vdots \\ \beta_{d0} + \sum_{k=1}^M \beta_{dk} \Psi(\gamma_{k0} + y_j^T \gamma_k) \end{pmatrix} &= \mu_{\xi_{2j}} + \epsilon_j & j = 1, \dots, m & \quad \epsilon_j \sim f_2 \in \mathbb{R}^d. \end{aligned}$$

Se observa que cada y_j se transforma en otro punto $\phi(y_j)$ cuyas coordenadas son una expresión no lineal expresada en términos de una combinación lineal de funciones logísticas. Estas funciones logísticas que se utilizan en el modelo son las mismas en todas las coordenadas, cambiando, sin embargo, los coeficientes que las acompañan en cada coordenada. Esta suposición deberá modificarse en el caso de $M = 1$ nodo por la existencia de problemas de invertibilidad y será tratado en la sección 4.3.

En forma matricial el modelo tiene la expresión,

$$\begin{aligned} x_i &= \mu_{\xi_{1i}} + \varepsilon_i & i = 1, \dots, n & \quad \varepsilon_i \sim f_1 \in \mathbb{R}^d \\ \beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) &= \mu_{\xi_{2j}} + \epsilon_j & i = 1, \dots, m & \quad \epsilon_j \sim f_2 \in \mathbb{R}^d, \end{aligned} \quad (4.1)$$

donde

$$\beta_0 = \begin{pmatrix} \beta_{10} \\ \vdots \\ \beta_{d0} \end{pmatrix}_{d \times 1} \quad B = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1M} \\ \vdots & & \vdots \\ \beta_{d1} & \cdots & \beta_{dM} \end{pmatrix}_{d \times M} \quad \Psi(\gamma_0 + y_j^T \gamma) = \begin{pmatrix} \Psi(\gamma_{10} + y_j^T \gamma_1) \\ \vdots \\ \Psi(\gamma_{M0} + y_j^T \gamma_M) \end{pmatrix}_{M \times 1} \quad (4.2)$$

La definición y desarrollo del Proceso de Poisson de tasa λ , $\{\mu_l\}$ $l = 1, \dots, N$, que genera las configuraciones y que está definido sobre una región $V \subset \mathbb{R}^d$ de volumen v , se puede encontrar en la sección 3.3.

Con respecto a la matriz con la que se definen los emparejamientos, recordamos que en el caso de dos configuraciones viene dada por

$$M_{jk} = \begin{cases} 1 & \text{si } \xi_{1j} = \xi_{2k} \\ 0 & \text{si } \xi_{1j} \neq \xi_{2k} \end{cases}, \quad j = 1, \dots, n, k = 1, \dots, m, \quad (4.3)$$

es decir, M_{jk} indica si x_j e y_k provienen del mismo punto μ_l y, por lo tanto, están emparejados.

Asumiendo que, conocido el número de emparejamientos entre las dos configuraciones L la distribución de \mathbf{M} es la uniforme, se obtiene que la distribución a priori de esta matriz de emparejamientos viene dada por

$$p(\mathbf{M}) \propto (\rho/\lambda v)^L,$$

donde ρ es la medida de la tendencia a priori de los puntos a estar emparejados (sección 2.4.3).

4.2.2. Construcción de la verosimilitud de los datos

De la misma manera que hicimos en 3.5.1, vamos a construir la expresión de la verosimilitud de los datos asumiendo conocida la matriz de emparejamientos,

diferenciando entre los puntos que están emparejados y los que no.

Así, de (4.1) se tiene que

$$\begin{aligned} f(x_i) &= f_1(x_i - \mu_{\xi_{1i}}) & i &= 1, \dots, n \\ \beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu_{\xi_{2j}} &\sim f_2 & j &= 1, \dots, m. \end{aligned}$$

Vamos a calcular la función de densidad de cada punto y_j $j = 1, \dots, m$ de la configuración \mathbf{y} .

$$f(y_j) = f_2(\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu_{\xi_{2j}}) |J_j|,$$

donde

$$\begin{aligned} J_j &= \frac{\partial}{\partial y_j} \left(\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu_{\xi_{2j}} \right) = B \frac{\partial}{\partial y_j} \left(\Psi(\gamma_0 + y_j^T \gamma) \right) \\ &= B \left(\frac{\partial}{\partial y_{js}} \Psi(\gamma_{k0} + y_j^T \gamma_k) \right)_{\substack{k=1, \dots, M \\ s=1, \dots, d}} \\ &= B \begin{pmatrix} \frac{\partial}{\partial y_{j1}} \Psi(\gamma_{10} + y_j^T \gamma_1) & \cdots & \frac{\partial}{\partial y_{jd}} \Psi(\gamma_{10} + y_j^T \gamma_1) \\ \vdots & & \vdots \\ \frac{\partial}{\partial y_{j1}} \Psi(\gamma_{M0} + y_j^T \gamma_M) & \cdots & \frac{\partial}{\partial y_{jd}} \Psi(\gamma_{M0} + y_j^T \gamma_M) \end{pmatrix}. \end{aligned}$$

Calculamos, fijado $j = 1, \dots, m$, y para todo $k = 1, \dots, M$ y $s = 1, \dots, d$,

$$\begin{aligned} \frac{\partial}{\partial y_{js}} \Psi(\gamma_{k0} + y_j^T \gamma_k) &= \frac{\partial}{\partial y_{js}} \left[\frac{\exp \left\{ \gamma_{k0} + \sum_{l=1}^d y_{jl} \gamma_{kl} \right\}}{1 + \exp \left\{ \gamma_{k0} + \sum_{l=1}^d y_{jl} \gamma_{kl} \right\}} \right] \\ &= \frac{\exp \left\{ \gamma_{k0} + \sum_{l=1}^d y_{jl} \gamma_{kl} \right\} \gamma_{ks}}{\left(1 + \exp \left\{ \gamma_{k0} + \sum_{l=1}^d y_{jl} \gamma_{kl} \right\} \right)^2} \\ &= \frac{\Psi(\gamma_{k0} + \sum_{l=1}^d y_{jl} \gamma_{kl}) \gamma_{ks}}{\left(1 + \exp \left\{ \gamma_{k0} + \sum_{l=1}^d y_{jl} \gamma_{kl} \right\} \right)} \\ &= \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k) \gamma_{ks}}{\left(1 + \exp \left\{ \gamma_{k0} + y_j^T \gamma_k \right\} \right)}. \end{aligned}$$

Por lo tanto,

$$\begin{aligned} \frac{\partial}{\partial y_j} (\Psi(\gamma_0 + y_j^T \gamma)) &= \begin{pmatrix} \frac{\Psi(\gamma_{10} + y_j^T \gamma_1)}{(1 + \exp\{\gamma_{10} + y_j^T \gamma_1\})} \gamma_1^T \\ \vdots \\ \frac{\Psi(\gamma_{M0} + y_j^T \gamma_M)}{(1 + \exp\{\gamma_{M0} + y_j^T \gamma_M\})} \gamma_M^T \end{pmatrix}_{M \times d} \\ &= \begin{pmatrix} \frac{\Psi(\gamma_{10} + y_j^T \gamma_1)}{(1 + \exp\{\gamma_{10} + y_j^T \gamma_1\})} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \frac{\Psi(\gamma_{M0} + y_j^T \gamma_M)}{(1 + \exp\{\gamma_{M0} + y_j^T \gamma_M\})} \end{pmatrix} \begin{pmatrix} \gamma_1^T \\ \vdots \\ \gamma_M^T \end{pmatrix}. \end{aligned}$$

Entonces, denotando por

$$\gamma = \begin{pmatrix} \gamma_1^T \\ \vdots \\ \gamma_M^T \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1d} \\ \vdots & & \vdots \\ \gamma_{M1} & \cdots & \gamma_{Md} \end{pmatrix}, \quad (4.4)$$

la expresión del jacobiano será

$$|J_j| = |B\gamma| \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp\{\gamma_{k0} + y_j^T \gamma_k\})},$$

donde $|B\gamma|$ es el valor absoluto del determinante de la matriz $B\gamma$ de dimensión $d \times d$.

Con este desarrollo hemos calculado las expresiones de las funciones de densidad de los puntos de las dos configuraciones en las que nos apoyaremos para calcular la verosimilitud de los datos y que son:

$$\begin{aligned} f(x_i) &= f_1(x_i - \mu_{\xi_{1i}}), \quad i = 1, \dots, n \\ f(y_j) &= f_2(\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu_{\xi_{2j}}) |B\gamma| \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp\{\gamma_{k0} + y_j^T \gamma_k\})}, \\ \text{para } j &= 1, \dots, m. \end{aligned} \quad (4.5)$$

A partir de (4.5), se va diferenciando si los puntos están o no emparejados para calcular la aportación de cada subgrupo de puntos a la verosimilitud total. Al mismo tiempo, se considera la región $V \subset \mathbb{R}^d$ suficientemente grande con respecto al soporte de f_1 y f_2 para aproximar V por \mathbb{R}^d , simplificándose así las expresiones.

Aportación a la verosimilitud de los puntos no emparejados de \mathbf{x}

De la misma manera que en el caso lineal (sección 3.5.1) se obtiene que para todo x_i , punto no emparejado de \mathbf{x} ,

$$f(x_i) = \frac{1}{v} \int_V f_1(x_i - \mu) d\mu,$$

por lo que la contribución conjunta de todos ellos será

$$\left(\frac{1}{v}\right)^{n-L} \prod_{\{i: M_{ij}=0 \forall j\}} \int_V f_1(x_i - \mu) d\mu.$$

Esta expresión puede aproximarse, sustituyendo V por \mathbb{R}^d , a

$$\left(\frac{1}{v}\right)^{n-L}. \quad (4.6)$$

Aportación a la verosimilitud de los puntos no emparejados de \mathbf{y}

Análogamente, y para todo y_j punto no emparejado de \mathbf{y} , nos basamos en la expresión (4.5). Así,

$$\begin{aligned} f(y_j) &= \frac{1}{v} \int_V f(y_j | \mu) d\mu \\ &= \frac{1}{v} |B\gamma| \int_V f_2(\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu) d\mu \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp \{ \gamma_{k0} + y_j^T \gamma_k \})}, \end{aligned}$$

por lo que la contribución conjunta de todos ellos se puede expresar como

$$\begin{aligned} &\left(\frac{1}{v}\right)^{m-L} |B\gamma|^{m-L} \prod_{\{j: M_{ij}=0 \forall i\}} \left[\int_V f_2(\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu) d\mu \right. \\ &\quad \left. \cdot \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp \{ \gamma_{k0} + y_j^T \gamma_k \})} \right]. \end{aligned}$$

De nuevo, esta expresión queda simplificada con la aproximación de V por \mathbb{R}^d ya que $\int_{\mathbb{R}^d} f_2(\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu) d\mu = 1$.

Así, la aportación aproximada a la verosimilitud de los puntos no emparejados de \mathbf{y} , tiene la forma,

$$\left(\frac{1}{v}\right)^{m-L} |B\gamma|^{m-L} \prod_{\{j: M_{ij}=0 \forall i\}} \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp \{\gamma_{k0} + y_j^T \gamma_k\})}. \quad (4.7)$$

Aportación a la verosimilitud de los puntos (x_i, y_j) emparejados

Por último, nos centramos en aquellos puntos x_i e y_j tales que $M_{ij} = 1$. De la misma manera que en 3.5.1 procedemos al cálculo de la expresión de $f(x_i, y_j)$ con $M_{ij} = 1$, basándonos de nuevo en las expresiones de (4.5). Así,

$$\begin{aligned} f(x_i, y_j) &= \frac{1}{v} \int_V f(x_i, y_j) d\mu = \frac{1}{v} \int_V f(x_i) f(y_j) d\mu \\ &= \frac{1}{v} |B\gamma| \int_V f_1(x_i - \mu) f_2(\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu) d\mu \\ &\quad \cdot \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_i^T \gamma_k)}{(1 + \exp \{\gamma_{k0} + y_i^T \gamma_k\})}, \end{aligned}$$

por lo que la contribución conjunta de todos ellos se puede expresar como

$$\begin{aligned} \left(\frac{1}{v} |B\gamma|\right)^L &\prod_{\{(i,j): M_{ij}=1\}} \left[\int_V f_1(x_i - \mu) f_2(\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu) d\mu \right. \\ &\quad \cdot \left. \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp \{\gamma_{k0} + y_j^T \gamma_k\})} \right]. \end{aligned}$$

De nuevo, construimos su expresión aproximada. Utilizando que la función de densidad de la variable $Z = \varepsilon_i - \epsilon_j$, donde $\varepsilon_i \sim f_1$ y $\epsilon_j \sim f_2$, viene dada en (3.13) por la expresión $g_{i,j}(z) = g(z) = \int_{\mathbf{R}^d} f_1(w) f_2(w - z) dw$, e identificando en nuestro caso $x_i - \mu = w$ y $\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu = w - z$, se tiene que

$$\int_{\mathbf{R}^d} f_1(x_i - \mu) f_2(\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu) d\mu = g(x_i - \beta_0 - B\Psi(\gamma_0 + y_j^T \gamma)).$$

Por tanto, la aportación aproximada a la verosimilitud de los puntos emparejados

toma la expresión,

$$\left(\frac{1}{v} |B\gamma|\right)^L \prod_{\{(i,j): M_{ij}=1\}} \left[g(x_i - \beta_0 - B\Psi(\gamma_0 + y_j^T \gamma)) \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp \{\gamma_{k0} + y_j^T \gamma_k\})} \right]. \quad (4.8)$$

En conclusión, el producto de (4.6), (4.7) y (4.8) da lugar a la expresión aproximada de la verosimilitud de los datos

$$\begin{aligned} p(\mathbf{x}, \mathbf{y} \mid M, \beta_0, B, \gamma_0, \gamma) &\approx \left(\frac{1}{v}\right)^{n+m-L} |B\gamma|^m \cdot \\ &\prod_{\{j: M_{ij}=0 \forall i\}} \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp \{\gamma_{k0} + y_j^T \gamma_k\})} \cdot \\ &\prod_{\{(i,j): M_{ij}=1\}} \left[g(x_i - \beta_0 - B\Psi(\gamma_0 + y_j^T \gamma)) \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp \{\gamma_{k0} + y_j^T \gamma_k\})} \right], \end{aligned} \quad (4.9)$$

donde $g(z)$ es la densidad de la diferencia de los errores $\varepsilon_i - \epsilon_j$.

Verosimilitud suponiendo errores normales

Si se asume que los errores se distribuyen como una normal, (4.1) se expresa como:

$$\begin{aligned} x_i &= \mu_{\xi_{1i}} + \varepsilon_i & i &= 1, \dots, n & \varepsilon_i &\sim \mathbf{N}_d(0, \sigma^2 \mathbf{I}_d) \\ \beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) &= \mu_{\xi_{2j}} + \epsilon_j & j &= 1, \dots, m & \epsilon_j &\sim \mathbf{N}_d(0, \sigma^2 \mathbf{I}_d). \end{aligned}$$

Para calcular qué expresión tiene la verosimilitud (4.9) bajo esta hipótesis, utilizamos el resultado (i) de la Proposición 5, donde se demuestra que

$$g(z) = \left(\frac{1}{\sigma\sqrt{2}}\right)^d \varphi_d\left(\frac{z}{\sigma\sqrt{2}}\right),$$

siendo $\varphi_d(\cdot)$ la función de densidad de la distribución $\mathbf{N}_d(0, \mathbf{I}_d)$. Entonces,

$$\begin{aligned}
 p(\mathbf{x}, \mathbf{y} \mid \mathbf{M}, \beta_0, B, \gamma_0, \gamma, \sigma^2) &\approx \left(\frac{1}{v}\right)^{n+m-L} |B\gamma|^m \cdot \\
 &\prod_{\{j:M_{ij}=0 \forall i\}} \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp\{\gamma_{k0} + y_j^T \gamma_k\})} \cdot \\
 &\prod_{\{(i,j):M_{ij}=1\}} \left[\left(\frac{1}{\sigma\sqrt{2}}\right)^d \varphi_d\left(\frac{x_i - \beta_0 - B\Psi(\gamma_0 + y_j^T \gamma)}{\sigma\sqrt{2}}\right) \cdot \right. \\
 &\left. \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp\{\gamma_{k0} + y_j^T \gamma_k\})} \right]. \tag{4.10}
 \end{aligned}$$

Desarrollamos la expresión de $\varphi_d(\frac{x_i - \beta_0 - B\Psi(\gamma_0 + y_j^T \gamma)}{\sigma\sqrt{2}})$ para sustituirla en (4.10).

$$\begin{aligned}
 \varphi_d\left(\frac{x_i - \beta_0 - B\Psi(\gamma_0 + y_j^T \gamma)}{\sigma\sqrt{2}}\right) &= \\
 &= \frac{1}{(\sqrt{2\pi})^d} \exp\left\{-\frac{1}{2} \left\| \frac{x_i - \beta_0 - B\Psi(\gamma_0 + y_j^T \gamma)}{\sigma\sqrt{2}} \right\|^2\right\} \\
 &= \frac{1}{(\sqrt{2\pi})^d} \exp\left\{-\frac{1}{4\sigma^2} \|x_i - \beta_0 - B\Psi(\gamma_0 + y_j^T \gamma)\|^2\right\} \\
 &= \frac{1}{(\sqrt{2\pi})^d} \exp\left\{\frac{1}{4\sigma^2} \sum_{l=1}^d \left(x_{il} - \beta_{l0} - \sum_{k=1}^M \beta_{lk} \Psi(\gamma_{k0} + y_j^T \gamma_k)\right)^2\right\}.
 \end{aligned}$$

Por lo tanto, la verosimilitud aproximada de los datos suponiendo que los errores se distribuyen según una normal es,

$$\begin{aligned}
 p(\mathbf{x}, \mathbf{y} \mid \mathbf{M}, \beta_0, B, \gamma_0, \gamma, \sigma^2) &\approx \left(\frac{1}{v}\right)^{n+m-L} |B\gamma|^m \cdot \left(\frac{1}{\sigma\sqrt{2}}\right)^{dL} \frac{1}{(\sqrt{2\pi})^{Ld}} \\
 &\prod_{\{j:M_{ij}=0 \forall i\}} \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp\{\gamma_{k0} + y_j^T \gamma_k\})} \cdot \\
 &\prod_{\{(i,j):M_{ij}=1\}} \left[\exp\left\{-\frac{1}{4\sigma^2} \sum_{l=1}^d \left(x_{il} - \beta_{l0} - \sum_{k=1}^M \beta_{lk} \Psi(\gamma_{k0} + y_j^T \gamma_k)\right)^2\right\} \cdot \right. \\
 &\left. \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp\{\gamma_{k0} + y_j^T \gamma_k\})} \right]. \tag{4.11}
 \end{aligned}$$

4.2.3. Inferencia bayesiana en el caso de configuraciones etiquetadas

Una vez que hemos desarrollado los resultados en la situación general con emparejamientos desconocidos entre las configuraciones, vamos a realizar inferencias sobre los parámetros suponiendo que las dos configuraciones tienen el mismo número de puntos (n), y que además se sabe que están todos emparejados, es decir, $L = n$.

Sin pérdida de generalidad, supongamos que $\forall i = 1, \dots, n$ los puntos x_i e y_i forman los n emparejamientos, es decir, $\xi_{1i} = \xi_{2i}$. Así,

$$\mathbf{M}_{jk} = \begin{cases} 1 & j = k \\ 0 & \text{resto} \end{cases}.$$

Bajo estas hipótesis el problema se centra en la estimación de los parámetros que definen la red neuronal, además de la varianza.

Siguiendo bajo la hipótesis de normalidad y prescindiendo de las constantes, (4.11) particularizada en configuraciones etiquetadas, tiene la forma

$$\begin{aligned} p(\mathbf{x}, \mathbf{y} \mid \beta_0, B, \gamma_0, \boldsymbol{\gamma}, \sigma^2) &\propto |B\boldsymbol{\gamma}|^n \left(\frac{1}{\sigma}\right)^{dn} \cdot \\ &\prod_{i=1}^n \left[\exp \left\{ -\frac{1}{4\sigma^2} \sum_{l=1}^d \left(x_{il} - \beta_{l0} - \sum_{k=1}^M \beta_{lk} \Psi(\gamma_{k0} + y_i^T \gamma_k) \right)^2 \right\} \cdot \right. \\ &\quad \left. \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_i^T \gamma_k)}{(1 + \exp \{ \gamma_{k0} + y_i^T \gamma_k \})} \right] \\ &\propto |B\boldsymbol{\gamma}|^n \left(\frac{1}{\sigma}\right)^{dn} \exp \left\{ -\frac{1}{4\sigma^2} \sum_{i=1}^n \sum_{l=1}^d \left(x_{il} - \beta_{l0} - \sum_{k=1}^M \beta_{lk} \Psi(\gamma_{k0} + y_i^T \gamma_k) \right)^2 \right\} \cdot \\ &\quad \prod_{i=1}^n \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_i^T \gamma_k)}{(1 + \exp \{ \gamma_{k0} + y_i^T \gamma_k \})}. \end{aligned}$$

Si nos centramos en \mathbb{R}^2 , la verosimilitud en el caso de dos configuraciones eti-

quetadas de n puntos, con todos ellos emparejados, puede reescribirse como

$$\begin{aligned}
 p(\mathbf{x}, \mathbf{y} \mid \beta_0, B, \gamma_0, \gamma, \sigma^2) &\propto |B\gamma|^n \left(\frac{1}{\sigma}\right)^{2n} \cdot \\
 &\exp \left\{ -\frac{1}{4\sigma^2} \left[\sum_{i=1}^n \left(x_{i1} - \beta_{10} - \sum_{k=1}^M \beta_{1k} \Psi(\gamma_{k0} + y_i^T \gamma_k) \right)^2 \right. \right. \\
 &\quad \left. \left. + \sum_{i=1}^n \left(x_{i2} - \beta_{20} - \sum_{k=1}^M \beta_{2k} \Psi(\gamma_{k0} + y_i^T \gamma_k) \right)^2 \right] \right\} \cdot \\
 &\prod_{i=1}^n \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_i^T \gamma_k)}{(1 + \exp \{ \gamma_{k0} + y_i^T \gamma_k \})}. \tag{4.12}
 \end{aligned}$$

A continuación se realizan inferencias sobre los parámetros del modelo, calculando las expresiones de las distribuciones a posteriori. Pero antes, dado que la verosimilitud depende de $|B\gamma|$, va a ser necesario desarrollar este determinante en función de los distintos parámetros para luego encontrar la expresión de las distribuciones a posteriori.

$$\begin{aligned}
 B\gamma &= \begin{pmatrix} \beta_{11} & \cdots & \beta_{1M} \\ \beta_{21} & \cdots & \beta_{2M} \end{pmatrix} \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \vdots & \vdots \\ \gamma_{M1} & \gamma_{M2} \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^M \beta_{1k} \gamma_{k1} & \sum_{k=1}^M \beta_{1k} \gamma_{k2} \\ \sum_{k=1}^M \beta_{2k} \gamma_{k1} & \sum_{k=1}^M \beta_{2k} \gamma_{k2} \end{pmatrix} \\
 &= \begin{pmatrix} \beta_1^T \gamma_{\cdot 1} & \beta_1^T \gamma_{\cdot 2} \\ \beta_2^T \gamma_{\cdot 1} & \beta_2^T \gamma_{\cdot 2} \end{pmatrix},
 \end{aligned}$$

donde $\beta_r^T = (\beta_{r1}, \dots, \beta_{rM})$ y $\gamma_{\cdot r} = (\gamma_{1r}, \dots, \gamma_{Mr})^T$ con $r = 1, 2$.

Por lo tanto,

$$|B\gamma| = |\beta_1^T \gamma_{\cdot 1} \beta_2^T \gamma_{\cdot 2} - \beta_2^T \gamma_{\cdot 1} \beta_1^T \gamma_{\cdot 2}|, \tag{4.13}$$

expresión que se desarrollará en términos de aquel parámetro del que estemos haciendo inferencias.

Distribución a posteriori de la varianza σ^2

De nuevo, asumimos como distribución a priori de la varianza σ^2 una gamma inversa $\sigma^{-2} \sim G(\alpha, \beta)$, $\alpha, \beta > 0$, cuya función de densidad es,

$$p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-(\alpha+1)} \exp\{-\beta/\sigma^2\}.$$

Entonces a partir de (4.12), e incluyendo en la constante de proporcionalidad todo aquello que no depende de σ^2 , la distribución a posteriori tendrá la forma:

$$\begin{aligned} p(\sigma^2 \mid \mathbf{x}, \mathbf{y}, \beta_0, B, \gamma_0, \gamma) &\propto (\sigma^2)^{-(\alpha+1)-n} \exp\{-\beta/\sigma^2\} \cdot \\ &\exp \left\{ -\frac{1}{4\sigma^2} \left[\sum_{i=1}^n \left(x_{i1} - \beta_{10} - \sum_{k=1}^M \beta_{1k} \Psi(\gamma_{k0} + y_i^T \gamma) \right)^2 \right. \right. \\ &\left. \left. + \sum_{i=1}^n \left(x_{i2} - \beta_{20} - \sum_{k=1}^M \beta_{2k} \Psi(\gamma_{k0} + y_i^T \gamma) \right)^2 \right] \right\}. \end{aligned}$$

Entonces, llamando

$$\begin{aligned} A = & \frac{1}{4} \sum_{i=1}^n \left(\left[x_{i1} - \beta_{10} - \sum_{k=1}^M \beta_{1k} \Psi(\gamma_{k0} + y_i^T \gamma_k) \right]^2 \right. \\ & \left. + \left[x_{i2} - \beta_{20} - \sum_{k=1}^M \beta_{2k} \Psi(\gamma_{k0} + y_i^T \gamma_k) \right]^2 \right), \end{aligned}$$

se llega a que la distribución a posteriori es $\sigma^{-2} \sim G(\alpha+n, \beta+A)$, ya que su función de densidad tiene la forma

$$p(\sigma^2 \mid \mathbf{x}, \mathbf{y}, \beta_0, B, \gamma_0, \gamma) \propto (\sigma^2)^{-(\alpha+1)-n} \exp\left\{-\frac{1}{\sigma^2}(\beta+A)\right\}.$$

Distribución a posteriori de β_{r0} , $r = 1, 2$.

Asumimos que la distribución a priori para $\beta_{r0} \sim N(\mu_{\beta_{r0}}, \sigma_{\beta_{r0}}^2)$. Con el objetivo de clarificar los cálculos, nos centramos en β_{10} .

Basándonos de nuevo en la verosimilitud (4.12), la distribución a posteriori de β_{10} tendrá la forma,

$$p(\beta_{10} \mid \mathbf{x}, \mathbf{y}, \beta_{20}, B, \gamma_0, \gamma, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma_{\beta_{10}}^2} (\beta_{10}^2 - 2\mu_{\beta_{10}}\beta_{10}) \right\} \cdot \exp \left\{ -\frac{1}{4\sigma^2} \sum_{i=1}^n (x_{i1} - \beta_{10} - \sum_{k=1}^M \beta_{1k} \Psi(\gamma_{k0} + y_i^T \gamma_k))^2 \right\}. \quad (4.14)$$

Identificando en la expresión lo que es independiente de β_{10} , para todo $i = 1, \dots, n$ se define

$$R_{i10} = x_{i1} - \sum_{k=1}^M \beta_{1k} \Psi(\gamma_{k0} + y_i^T \gamma_k), \quad (4.15)$$

por lo que (4.14) se puede escribir de la forma

$$p(\beta_{10} \mid \mathbf{x}, \mathbf{y}, \beta_{20}, B, \gamma_0, \gamma, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma_{\beta_{10}}^2} (\beta_{10}^2 - 2\mu_{\beta_{10}}\beta_{10}) - \frac{1}{4\sigma^2} \sum_{i=1}^n (R_{i10} - \beta_{10})^2 \right\}.$$

Desarrollando el cuadrado e incluyendo en la constante de proporcionalidad aquello que no depende de β_{10} , se llega a que

$$\begin{aligned} p(\beta_{10} \mid \mathbf{x}, \mathbf{y}, \beta_{20}, B, \gamma_0, \gamma, \sigma^2) &\propto \exp \left\{ -\frac{1}{2\sigma_{\beta_{10}}^2} (\beta_{10}^2 - 2\mu_{\beta_{10}}\beta_{10}) \right. \\ &\quad \left. - \frac{1}{4\sigma^2} (n\beta_{10}^2 - 2\beta_{10} \sum_{i=1}^n R_{i10}) \right\} \\ &= \exp \left\{ -\frac{1}{2} \left[\beta_{10}^2 \left(\frac{1}{\sigma_{\beta_{10}}^2} + \frac{n}{2\sigma^2} \right) - \beta_{10} \left(\frac{2\mu_{\beta_{10}}}{\sigma_{\beta_{10}}^2} + \frac{\sum_{i=1}^n R_{i10}}{\sigma^2} \right) \right] \right\} \\ &= \exp \left\{ -\frac{1}{2} [\beta_{10}^2 C_{10} - \beta_{10} D_{10}] \right\}, \end{aligned}$$

donde C_{10} y D_{10} son las expresiones independientes de β_{10} ,

$$\begin{aligned} C_{10} &= \frac{1}{\sigma_{\beta_{10}}^2} + \frac{n}{2\sigma^2} \\ D_{10} &= \frac{2\mu_{\beta_{10}}}{\sigma_{\beta_{10}}^2} + \frac{\sum_{i=1}^n R_{i10}}{\sigma^2}. \end{aligned} \quad (4.16)$$

Completando cuadrados, se llega a que la expresión de la distribución a posteriori de β_{10} tiene la forma

$$p(\beta_{10} \mid \mathbf{x}, \mathbf{y}, \beta_{20}, B, \gamma_0, \boldsymbol{\gamma}, \sigma^2) \propto \exp \left\{ -\frac{1}{2C_{10}^{-1}} \left(\beta_{10} - \frac{D_{10}}{2C_{10}} \right)^2 \right\},$$

que identifica a la distribución normal

$$\beta_{10} \mid \mathbf{x}, \mathbf{y}, \beta_{20}, B, \gamma_0, \boldsymbol{\gamma}, \sigma^2 \sim N\left(\frac{D_{10}}{2C_{10}}, C_{10}^{-1}\right),$$

donde C_{10} y D_{10} vienen definidas en (4.16).

Un resultado totalmente análogo se obtiene para β_{20} por lo que, para $r = 1, 2$, y denotando por $\beta_{0(r)}$ el vector β_0 sin la coordenada r , la distribución a posteriori de β_{r0} es

$$\beta_{r0} \mid \mathbf{x}, \mathbf{y}, \beta_{0(r)}, B, \gamma_0, \boldsymbol{\gamma}, \sigma^2 \sim N\left(\frac{D_{r0}}{2C_{r0}}, C_{r0}^{-1}\right),$$

donde

$$\begin{aligned} C_{r0} &= \frac{1}{\sigma_{\beta_{r0}}^2} + \frac{n}{2\sigma^2}, \\ D_{r0} &= \frac{2\mu_{\beta_{r0}}}{\sigma_{\beta_{r0}}^2} + \frac{\sum_{i=1}^n R_{ir0}}{\sigma^2}, \end{aligned}$$

siendo para todo $i = 1, \dots, n$,

$$R_{ir0} = x_{ir} - \sum_{k=1}^M \beta_{jk} \Psi(\gamma_{k0} + y_i^T \gamma_k).$$

Distribución a posteriori de β_{rs} $r = 1, 2$, $s = 1, \dots, M$

Los parámetros β_{rs} representan los coeficientes de las combinaciones lineales de las funciones logísticas para cada una de las dos coordenadas. No habiendo inicialmente ninguna restricción sobre ellos, asumimos que a priori siguen una distribución

$N(\mu_{\beta_{rs}}, \sigma_{\beta_{rs}}^2)$. Denotamos por $B_{(rs)}$ la matriz B , que recoge todos los parámetros betas, menos el β_{rs} . Así,

$$p(\beta_{rs} \mid \mathbf{x}, \mathbf{y}, \beta_0, B_{(rs)}, \gamma_0, \gamma, \sigma^2) \propto |B\gamma|^n \exp\left\{-\frac{1}{2\sigma_{\beta_{rs}}^2}(\beta_{rs}^2 - 2\mu_{\beta_{rs}}\beta_{rs})\right\} \cdot \\ \exp\left\{-\frac{1}{4\sigma^2} \left[\sum_{i=1}^n \left(x_{i1} - \beta_{10} - \sum_{k=1}^M \beta_{1k} \Psi(\gamma_{k0} + y_i^T \gamma) \right)^2 \right. \right. \\ \left. \left. + \sum_{i=1}^n \left(x_{i2} - \beta_{20} - \sum_{k=1}^M \beta_{2k} \Psi(\gamma_{k0} + y_i^T \gamma) \right)^2 \right] \right\}.$$

De nuevo y sin pérdida de generalidad, nos centramos en el caso $r = 1$. Identificando con R_{i1s} al sumando i de la exponencial que es independiente de β_{1s} ,

$$R_{i1s} = x_{i1} - \beta_{10} - \sum_{\substack{k=1 \\ k \neq s}}^M \beta_{1k} \Psi(\gamma_{k0} + y_i^T \gamma_k),$$

la distribución a posteriori de β_{1s} se podrá expresar como,

$$p(\beta_{1s} \mid \mathbf{x}, \mathbf{y}, \beta_0, B_{(1s)}, \gamma_0, \gamma, \sigma^2) \propto |B\gamma|^n \cdot \\ \exp\left\{-\frac{1}{2\sigma_{\beta_{1s}}^2}(\beta_{1s}^2 - 2\mu_{\beta_{1s}}\beta_{1s}) - \frac{1}{4\sigma^2} \sum_{i=1}^n (R_{i1s} - \beta_{1s} \Psi(\gamma_{s0} + y_i^T \gamma_s))^2\right\}.$$

En primer lugar, desarrollamos el determinante $|B\gamma|$ para ver cómo se puede expresar en términos de β_{1s} .

Operando sobre (4.13),

$$\begin{aligned} |B\gamma| &= |\beta_1^T \gamma_{\cdot 1} \beta_2^T \gamma_{\cdot 2} - \beta_2^T \gamma_{\cdot 1} \beta_1^T \gamma_{\cdot 2}| = |\gamma_{\cdot 2}^T \beta_2 \gamma_{\cdot 1}^T \beta_1 - \beta_2^T \gamma_{\cdot 1} \gamma_{\cdot 2}^T \beta_1| \\ &= |(\gamma_{\cdot 2}^T \beta_2 \gamma_{\cdot 1}^T - \beta_2^T \gamma_{\cdot 1} \gamma_{\cdot 2}^T) \beta_1|. \end{aligned}$$

Denotamos por $Q_1^T = \gamma_{\cdot 2}^T \beta_2 \gamma_{\cdot 1}^T - \beta_2^T \gamma_{\cdot 1} \gamma_{\cdot 2}^T$ que es un vector M -dimensional. Entonces,

$$|B\gamma| = \left| \sum_{k=1}^M Q_{1k} \beta_{1k} \right| = \left| Q_{1s} \beta_{1s} + \sum_{k=1, k \neq s}^M Q_{1k} \beta_{1k} \right| = |Q_{1s} \beta_{1s} + k_{1s}|, \quad (4.17)$$

donde k_{1s} recoge la expresión que queda independiente de β_{1s} .

En segundo lugar, se desarrolla el cuadrado del exponente y se completa cuadrados, de la misma manera que se hacía para el parámetro β_{r0} . De esta manera, se llega a que

$$p(\beta_{1s} \mid \mathbf{x}, \mathbf{y}, \beta_0, B_{(1s)}, \gamma_0, \boldsymbol{\gamma}, \sigma^2) \propto |Q_{1s}\beta_{1s} + k_{1s}|^n \exp\left\{-\frac{1}{2C_{1s}^{-1}}\left(\beta_{1s} - \frac{D_{1s}}{2C_{1s}}\right)^2\right\},$$

donde

$$\begin{aligned} C_{1s} &= \frac{\sum_{i=1}^n \Psi^2(\gamma_{s0} + y_i^T \gamma_s)}{2\sigma^2} + \frac{1}{\sigma_{\beta_{1s}}^2}, \\ D_{1s} &= \frac{\sum_{i=1}^n R_{i1s} \Psi(\gamma_{s0} + x_i^T \gamma_s)}{\sigma^2} + \frac{2\mu_{\beta_{1s}}}{\sigma_{\beta_{1s}}^2}, \end{aligned}$$

y para todo $i = 1, \dots, n$,

$$R_{i1s} = x_{i1} - \beta_{10} - \sum_{\substack{k=1 \\ k \neq s}}^M \beta_{1k} \Psi(\gamma_{k0} + y_i^T \gamma_k).$$

Un resultado totalmente análogo se obtiene para β_{2s} cambiando únicamente la expresión (4.17) de $|B\boldsymbol{\gamma}|$ como función de β_{2s} . Así,

$$|B\boldsymbol{\gamma}| = \left| \sum_{k=1}^M Q_{2k} \beta_{1k} \right| = \left| Q_{2s} \beta_{2s} + \sum_{\substack{k=1 \\ k \neq s}}^M Q_{2k} \beta_{2k} \right| = |Q_{2s} \beta_{2s} + k_{2s}|,$$

donde, en este caso, $Q_2^T = \beta_1^T \gamma_{\cdot 1} \gamma_{\cdot 2}^T - \gamma_{\cdot 2}^T \beta_1 \gamma_{\cdot 1}^T$.

En conclusión, la distribución a posteriori para β_{rs} , $r = 1, 2$, $s = 1, \dots, M$ es

$$p(\beta_{rs} \mid \mathbf{x}, \mathbf{y}, \beta_0, B_{(rs)}, \gamma_0, \boldsymbol{\gamma}, \sigma^2) \propto |Q_{rs}\beta_{rs} + k_{rs}|^n \exp\left\{-\frac{C_{rs}}{2}\left(\beta_{rs} - \frac{D_{rs}}{2C_{rs}}\right)^2\right\},$$

donde

$$\begin{aligned}
 Q_1^T &= \gamma_{\cdot 2}^T \beta_2 \gamma_{\cdot 1}^T - \beta_2^T \gamma_{\cdot 1} \gamma_{\cdot 2}^T \\
 Q_2^T &= \beta_1^T \gamma_{\cdot 1} \gamma_{\cdot 2}^T - \gamma_{\cdot 2}^T \beta_1 \gamma_{\cdot 1}^T \\
 k_{rs} &= \sum_{\substack{k=1 \\ k \neq s}}^M Q_{rk} \beta_{rk}, \\
 C_{rs} &= \frac{\sum_{i=1}^n \Psi^2(\gamma_{s0} + y_i^T \gamma_s)}{2\sigma^2} + \frac{1}{\sigma_{\beta_{rs}}^2}, \\
 D_{rs} &= \frac{\sum_{i=1}^n R_{irs} \Psi(\gamma_{s0} + y_i^T \gamma_s)}{\sigma^2} + \frac{2\mu_{\beta_{rs}}}{\sigma_{\beta_{rs}}^2}, \text{ y}
 \end{aligned}$$

y para todo $i = 1, \dots, n$,

$$R_{irs} = x_{ir} - \beta_{r0} - \sum_{\substack{k=1 \\ k \neq s}}^M \beta_{rs} \Psi(\gamma_{k0} + y_i^T \gamma_k). \quad (4.18)$$

Distribución a posteriori de γ_{r0} , $r = 1, \dots, M$

Para los parámetros γ_{r0} también asumimos distribuciones normales a priori, es decir, $\gamma_{r0} \sim N(\mu_{\gamma_{r0}}, \sigma_{\gamma_{r0}}^2)$. Entonces basándonos en la verosimilitud (4.12), y denotando por $\gamma_{0(r)}$ el vector γ_0 sin la coordenada r ,

$$\begin{aligned}
 p(\gamma_{r0} \mid \mathbf{x}, \mathbf{y}, \beta_0, B, \boldsymbol{\gamma}, \gamma_{0(r)}, \sigma^2) &\propto \exp \left\{ -\frac{1}{2\sigma_{\gamma_{r0}}^2} (\gamma_{r0}^2 - 2\mu_{\gamma_{r0}} \gamma_{r0}) \right\} \\
 &\cdot \exp \left\{ -\frac{1}{4\sigma^2} \left[\sum_{i=1}^n \left(x_{i1} - \beta_{10} - \sum_{k=1}^M \beta_{1k} \Psi(\gamma_{k0} + y_i^T \gamma_k) \right)^2 \right. \right. \\
 &\left. \left. + \sum_{i=1}^n \left(x_{i2} - \beta_{20} - \sum_{k=1}^M \beta_{2k} \Psi(\gamma_{k0} + y_i^T \gamma_k) \right)^2 \right] \right\} \cdot \\
 &\prod_{i=1}^n \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_i^T \gamma_k)}{(1 + \exp \{ \gamma_{k0} + y_i^T \gamma_k \})}. \quad (4.19)
 \end{aligned}$$

La forma de operar sobre ella, para simplificarla y expresarla como función de γ_{r0} , no es única. En este caso, se ha procedido a buscar una función en términos de una exponencial, por lo que comenzamos operando sobre el último factor.

$$\begin{aligned}
 & \prod_{i=1}^n \prod_{k=1}^M \frac{\Psi(\gamma_{k0} + y_i^T \gamma_k)}{(1 + \exp \{\gamma_{k0} + y_i^T \gamma_k\})} = \\
 & = \frac{\prod_{i=1}^n \prod_{k=1}^M \{\exp \{\gamma_{k0} + y_i^T \gamma_k\} / (1 + \exp \{\gamma_{k0} + y_i^T \gamma_k\})\}}{\prod_{i=1}^n \prod_{k=1}^M (1 + \exp \{\gamma_{k0} + y_i^T \gamma_k\})} \\
 & = \frac{\exp \left\{ \sum_{i=1}^n \sum_{k=1}^M (\gamma_{k0} + y_i^T \gamma_k) \right\}}{\exp \left\{ L n \left(\prod_{i=1}^n \prod_{k=1}^M (1 + \exp \{\gamma_{k0} + y_i^T \gamma_k\})^2 \right) \right\}} \\
 & = \exp \left\{ \sum_{i=1}^n \sum_{k=1}^M (\gamma_{k0} + y_i^T \gamma_k) - 2 \sum_{i=1}^n \sum_{k=1}^M L n (1 + \exp \{\gamma_{k0} + y_i^T \gamma_k\}) \right\} \quad (4.20)
 \end{aligned}$$

Entonces, identificando aquellos términos que no dependen de γ_{r0} y operando en (4.19),

$$\begin{aligned}
 & p(\gamma_{r0} \mid \mathbf{x}, \mathbf{y}, \beta_0, B, \boldsymbol{\gamma}, \gamma_{0(r)}, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma_{\gamma_{r0}}^2} (\gamma_{r0}^2 - 2\mu_{\gamma_{r0}} \gamma_{r0}) \right\} \cdot \\
 & \exp \left\{ -\frac{1}{4\sigma^2} \left[\sum_{i=1}^n \left(x_{i1} - \beta_{10} - \beta_{1r} \Psi(\gamma_{r0} + y_i^T \gamma_r) - \sum_{\substack{k=1 \\ k \neq r}}^M \beta_{1k} \Psi(\gamma_{k0} + y_i^T \gamma_k) \right)^2 \right. \right. \\
 & \left. \left. + \sum_{i=1}^n \left(x_{i2} - \beta_{20} - \beta_{2r} \Psi(\gamma_{r0} + y_i^T \gamma_r) - \sum_{\substack{k=1 \\ k \neq r}}^M \beta_{2k} \Psi(\gamma_{k0} + y_i^T \gamma_k) \right)^2 \right] \right\} \cdot \\
 & \exp \left\{ n\gamma_{r0} + \sum_{i=1}^n y_i^T \gamma_r + \sum_{i=1}^n \sum_{\substack{k=1 \\ k \neq r}}^M (\gamma_{k0} + y_i^T \gamma_k) \right. \\
 & \left. - 2 \sum_{i=1}^n L n (1 + \exp \{\gamma_{r0} + y_i^T \gamma_r\}) - 2 \sum_{i=1}^n \sum_{\substack{k=1 \\ k \neq r}}^M L n (1 + \exp \{\gamma_{k0} + y_i^T \gamma_k\}) \right\} \cdot
 \end{aligned}$$

De nuevo, recurriendo a las definiciones de R_{irs} dadas en (4.18), desarrollando los cuadrados e introduciendo en la constante de proporcionalidad los términos que

no dependen de γ_{r0} ,

$$\begin{aligned}
 p(\gamma_{r0} \mid \mathbf{x}, \mathbf{y}, \beta_0, B, \boldsymbol{\gamma}, \sigma^2) &\propto \exp \left\{ -\frac{1}{2\sigma_{\gamma_{r0}}^2} (\gamma_{r0}^2 - 2\mu_{\gamma_{r0}} \gamma_{r0}) \right\} \cdot \\
 &\exp \left\{ -\frac{1}{4\sigma^2} \left[\beta_{1r}^2 \sum_{i=1}^n \Psi^2(\gamma_{r0} + y_i^T \gamma_r) - 2\beta_{1r} \sum_{i=1}^n R_{i1r} \Psi(\gamma_{r0} + y_i^T \gamma_r) \right. \right. \\
 &\left. \left. + \beta_{2r}^2 \sum_{i=1}^n \Psi^2(\gamma_{r0} + y_i^T \gamma_r) - 2\beta_{2r} \sum_{i=1}^n R_{i2r} \Psi(\gamma_{r0} + y_i^T \gamma_r) \right] \right\} \cdot \\
 &\exp \left(n\gamma_{r0} - 2 \sum_{i=1}^n \ln(1 + \exp \{ \gamma_{r0} + y_i^T \gamma_r \}) \right) \\
 &\propto \exp \left\{ -\frac{1}{2\sigma_{\gamma_{r0}}^2} (\gamma_{r0}^2 - 2\mu_{\gamma_{r0}} \gamma_{r0}) \right. \\
 &\left. - \frac{1}{4\sigma^2} \left[(\beta_{1r}^2 + \beta_{2r}^2) \sum_{i=1}^n \Psi^2(\gamma_{r0} + y_i^T \gamma_r) - 2\beta_{1r} \sum_{i=1}^n R_{i1r} \Psi(\gamma_{r0} + y_i^T \gamma_r) \right. \right. \\
 &\left. \left. - 2\beta_{2r} \sum_{i=1}^n R_{i2r} \Psi(\gamma_{r0} + y_i^T \gamma_r) \right] + n\gamma_{r0} - 2 \sum_{i=1}^n \ln(1 + \exp \{ \gamma_{r0} + y_i^T \gamma_r \}) \right\}.
 \end{aligned}$$

Por lo tanto la expresión de la distribución a posteriori de γ_{r0} $r = 1, 2$ es

$$\begin{aligned}
 p(\gamma_{r0} \mid \mathbf{x}, \mathbf{y}, \beta_0, B, \boldsymbol{\gamma}, \gamma_{0(r)}, \sigma^2) &\propto \exp \left\{ -\frac{1}{2\sigma_{\gamma_{r0}}^2} (\gamma_{r0}^2 - 2\mu_{\gamma_{r0}} \gamma_{r0}) \right. \\
 &\left. - \frac{1}{4\sigma^2} \left[\sum_{s=1}^2 \sum_{i=1}^n \beta_{sr}^2 \Psi^2(\gamma_{r0} + y_i^T \gamma_r) - 2 \sum_{s=1}^2 \sum_{i=1}^n \beta_{sr} R_{isr} \Psi(\gamma_{r0} + y_i^T \gamma_r) \right] \right. \\
 &\left. + n\gamma_{r0} - 2 \sum_{i=1}^n \ln(1 + \exp \{ \gamma_{r0} + y_i^T \gamma_r \}) \right\},
 \end{aligned}$$

donde para todo $i = 1, \dots, n$ y $s, r = 1, 2$ R_{isr} tiene la expresión dada en (4.18).

Distribución a posteriori de γ_{rs} $r = 1, \dots, M$ y $s = 1, 2$

Para los parámetros gamma, también se van a suponer distribuciones a priori normales, es decir $\gamma_{rs} \sim N(\mu_{\gamma_{rs}}, \sigma_{\gamma_{rs}}^2)$. Recordamos que estos parámetros son los que forman parte de las funciones logísticas en el modelo.

Volvemos a utilizar la notación $\gamma_{(rs)}$ representando a la matriz γ definida en (4.4) salvo el parámetro γ_{rs} . Entonces, utilizando la verosimilitud (4.12), la distribución a posteriori tendrá la forma,

$$\begin{aligned}
 p(\gamma_{rs} \mid \mathbf{x}, \mathbf{y}, \beta_0, B, \gamma_0, \gamma_{(rs)}, \sigma^2) &\propto |B\gamma|^n \exp \left\{ -\frac{1}{2\sigma_{\gamma_{rs}}^2} (\gamma_{rs}^2 - 2\mu_{\gamma_{rs}} \gamma_{rs}) \right\} \cdot \\
 &\exp \left\{ -\frac{1}{4\sigma^2} \left[\sum_{i=1}^n (x_{i1} - \beta_{10} - \sum_{\substack{k=1 \\ k \neq r}}^M \beta_{1k} \Psi(\gamma_{k0} + y_i^T \gamma_k) - \beta_{1r} \Psi(\gamma_{r0} + y_i^T \gamma_r))^2 \right. \right. \\
 &\quad \left. \left. + \sum_{i=1}^n (x_{i2} - \beta_{20} - \sum_{\substack{k=1 \\ k \neq r}}^M \beta_{2k} \Psi(\gamma_{k0} + y_i^T \gamma_k) - \beta_{2r} \Psi(\gamma_{r0} + y_i^T \gamma_r))^2 \right] \right\} \cdot \\
 &\exp \left\{ \sum_{i=1}^n \sum_{k=1}^M (\gamma_{k0} + y_i^T \gamma_k) - 2 \sum_{i=1}^n \sum_{k=1}^M \ln(1 + \exp \{ \gamma_{k0} + y_i^T \gamma_k \}) \right\} \\
 &\propto |B\gamma|^n \exp \left\{ -\frac{1}{2\sigma_{\gamma_{rs}}^2} (\gamma_{rs}^2 - 2\mu_{\gamma_{rs}} \gamma_{rs}) \right\} \cdot \\
 &\exp \left\{ -\frac{1}{4\sigma^2} \left[\sum_{i=1}^n (R_{i1r} - \beta_{1r} \Psi(\gamma_{r0} + y_i^T \gamma_r))^2 + \sum_{i=1}^n (R_{i2r} - \beta_{2r} \Psi(\gamma_{r0} + y_i^T \gamma_r))^2 \right] \right\} \cdot \\
 &\exp \left\{ \sum_{i=1}^n \sum_{k=1}^M (\gamma_{k0} + y_i^T \gamma_k) - 2 \sum_{i=1}^n \sum_{k=1}^M \ln(1 + \exp \{ \gamma_{k0} + y_i^T \gamma_k \}) \right\}, \quad (4.21)
 \end{aligned}$$

donde R_{irs} vienen definidos en (4.18).

Como en (4.17), desarrollamos el determinante $|B\gamma|$ a partir de (4.13) para expresarlo en términos de γ_{rs} . A esta expresión la denotamos con h_{rs} . Nos centramos primero en el caso $s = 1$, es decir, en el parámetro γ_{r1} .

Como $\beta_1^T \gamma_{\cdot 1} = \gamma_{\cdot 1}^T \beta_1$ y $\beta_2^T \gamma_{\cdot 1} = \gamma_{\cdot 1}^T \beta_2$,

$$\begin{aligned}
 h_{r1} &= |B\gamma| = |\gamma_{\cdot 1}^T \beta_1 \beta_2^T \gamma_{\cdot 2} - \gamma_{\cdot 1}^T \beta_2 \beta_1^T \gamma_{\cdot 2}| \\
 &= |\gamma_{\cdot 1}^T (\beta_1 \beta_2^T \gamma_{\cdot 2} - \beta_2 \beta_1^T \gamma_{\cdot 2})| \\
 &= \left| \left(\gamma_{r1} \beta_{1r} + \sum_{\substack{k=1 \\ k \neq r}}^M \gamma_{k1} \beta_{1k} \right) \beta_2^T \gamma_{\cdot 2} - \left(\gamma_{r1} \beta_{2r} + \sum_{\substack{k=1 \\ k \neq r}}^M \gamma_{k1} \beta_{2k} \right) \beta_1^T \gamma_{\cdot 2} \right| \\
 &= |\gamma_{r1} \beta_{1r} \beta_2^T \gamma_{\cdot 2} - \gamma_{r1} \beta_{2r} \beta_1^T \gamma_{\cdot 2} + k k_{r1}|, \quad (4.22)
 \end{aligned}$$

donde

$$kk_{r1} = \sum_{\substack{k=1 \\ k \neq r}}^M \gamma_{k1} \beta_{1k} \beta_2^T \gamma_{\cdot 2} - \sum_{\substack{k=1 \\ k \neq r}}^M \gamma_{k1} \beta_{2k} \beta_1^T \gamma_{\cdot 2},$$

es el término que no depende de γ_{r1} .

Análogamente, en el caso de γ_{r2} , se llega a que

$$h_{r2} = |B\gamma| = |\gamma_{r2} \beta_{2r} \beta_1^T \gamma_{\cdot 1} - \gamma_{r2} \beta_{1r} \beta_2^T \gamma_{\cdot 1} + kk_{r2}|, \quad (4.23)$$

con

$$kk_{r2} = \sum_{\substack{k=1 \\ k \neq r}}^M \gamma_{k2} \beta_{2k} \beta_1^T \gamma_{\cdot 1} - \sum_{\substack{k=1 \\ k \neq r}}^M \gamma_{k2} \beta_{1k} \beta_2^T \gamma_{\cdot 1}.$$

Por otra parte, desarrollando cada una de las sumas de cuadrados de la exponencial en (4.21) y realizando operaciones de la misma manera que en el cálculo de la distribución a posteriori de γ_{r0} , se tiene que la expresión de la distribución a posteriori de γ_{rs} con $r = 1, \dots, M$ y $s = 1, 2$ es,

$$\begin{aligned} p(\gamma_{rs} \mid \mathbf{x}, \mathbf{y}, \beta_0, B, \gamma_0, \gamma_{(rs)}, \sigma^2) &\propto |h_{rs}|^n \exp \left\{ -\frac{1}{4\sigma^2} \sum_{s=1}^2 \sum_{i=1}^n \beta_{sr}^2 \Psi^2(\gamma_{r0} + y_i^T \gamma_r) \right. \\ &+ \frac{1}{2\sigma^2} \sum_{s=1}^2 \sum_{i=1}^n \beta_{sr} R_{isr} \Psi(\gamma_{r0} + y_i^T \gamma_r) + \gamma_{rs} \sum_{i=1}^n y_{is} \\ &\left. - 2 \sum_{i=1}^n \ln(1 + \exp \{ \gamma_{r0} + y_i^T \gamma_r \}) - \frac{1}{2\sigma_{\gamma_{rs}}^2} (\gamma_{rs}^2 - 2\mu_{\gamma_{rs}} \gamma_{rs}) \right\}, \end{aligned}$$

donde h_{rs} viene definido en (4.22) y en (4.23), y R_{isr} en (4.18), $i = 1, \dots, n$.

4.3. El modelo con $M = 1$ nodo, en el caso de \mathbb{R}^2

El modelo (4.1) con $M = 1$, y considerando las configuraciones representadas en dos dimensiones, tiene la expresión,

$$\begin{aligned} x_i &= \mu_{\xi_{1i}} + \varepsilon_i & i = 1, \dots, n & \quad \varepsilon_i \sim f_1 \in \mathbb{R}^2 \\ \begin{pmatrix} \beta_{10} + \beta_1 \Psi(\gamma_0 + y_j^T \gamma) \\ \beta_{20} + \beta_2 \Psi(\gamma_0 + y_j^T \gamma) \end{pmatrix} &= \mu_{\xi_{2j}} + \epsilon_j & i = 1, \dots, m & \quad \epsilon_j \sim f_2 \in \mathbb{R}^2, \end{aligned}$$

con ε_i y ϵ_j $i = 1, \dots, n$, $j = 1, \dots, m$ independientes, y donde

$$\Psi(\gamma_0 + y_j \gamma) = \frac{1}{\exp\{-(\gamma_0 + \gamma_1 y_{j1} + \gamma_2 y_{j2})\}}.$$

El problema que plantea este modelo es que la función no lineal $\phi : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$,

$$\phi(y_j) = \begin{pmatrix} \beta_{10} + \beta_1 \Psi(\gamma_0 + y_j^T \gamma) \\ \beta_{20} + \beta_2 \Psi(\gamma_0 + y_j^T \gamma) \end{pmatrix}$$

no es inyectiva, ya que existen $y_j \neq y_k$ tales que $\phi(y_j) = \phi(y_k)$ y, por lo tanto, no define una transformación invertible.

Siguiendo la notación matricial utilizada en (4.1), el modelo se puede expresar,

$$\begin{aligned} x_i &= \mu_{\xi_{1i}} + \varepsilon_i & i = 1, \dots, n & \quad \varepsilon_i \sim f_1 \in \mathbb{R}^2 \\ \beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) &= \mu_{\xi_{2j}} + \epsilon_j & j = 1, \dots, m & \quad \epsilon_j \sim f_2 \in \mathbb{R}^2, \end{aligned}$$

donde

$$\beta_0 = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} \quad B = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}.$$

Al calcular la función de densidad de los puntos y_j , $j = 1, \dots, m$, se llega a que el jacobiano es igual a cero, por lo que no es posible calcular la verosimilitud de los puntos y_j .

$$\begin{aligned} |J_j| &= \left| \frac{\partial}{\partial y_j} (\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu_{\xi_{2j}}) \right| = \left| B \frac{\partial}{\partial y_j} (\Psi(\gamma_0 + y_j^T \gamma)) \right| \\ &= \left| \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \left(\frac{\partial}{\partial y_{j1}} \Psi(\gamma_0 + y_j^T \gamma), \frac{\partial}{\partial y_{j2}} \Psi(\gamma_0 + y_j^T \gamma) \right) \right| \\ &= \left| \begin{pmatrix} \beta_1 \gamma_1 C_j & \beta_1 \gamma_2 C_j \\ \beta_2 \gamma_1 C_j & \beta_2 \gamma_2 C_j \end{pmatrix} \right| = C_j^2 (\beta_1 \gamma_1 \beta_2 \gamma_2 - \beta_2 \gamma_1 \beta_1 \gamma_2) = 0, \end{aligned}$$

con

$$C_j = \frac{\Psi(\gamma_0 + y_j^T \gamma)}{\exp(\gamma_0 + y_j^T \gamma)}.$$

Una alternativa que garantiza que la función no lineal sea biyectiva, es asumir que las combinaciones lineales que definen las funciones logísticas son distintas para

las dos coordenadas, viniendo definidas por los vectores $\gamma_0 = (\gamma_{10}, \gamma_{20})^T$, $\gamma_1 = (\gamma_{11}, \gamma_{12})^T$ y $\gamma_2 = (\gamma_{21}, \gamma_{22})^T$. Así, un modelo alternativo es:

$$\begin{aligned} x_i &= \mu_{\xi_{1i}} + \varepsilon_i & i = 1, \dots, n & \quad \varepsilon_i \sim f_1 \in \mathbb{R}^2 \\ \begin{pmatrix} \beta_{10} + \beta_1 \Psi(\gamma_{10} + y_j^T \gamma_1) \\ \beta_{20} + \beta_2 \Psi(\gamma_{20} + y_j^T \gamma_2) \end{pmatrix} &= \mu_{\xi_{2j}} + \epsilon_j & j = 1, \dots, m & \quad \epsilon_j \sim f_2 \in \mathbb{R}^2. \end{aligned}$$

Matricialmente,

$$\begin{aligned} x_i &= \mu_{\xi_{1i}} + \varepsilon_i & i = 1, \dots, n & \quad \varepsilon_i \sim f_1 \in \mathbb{R}^2 \\ \beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) &= \mu_{\xi_{2j}} + \epsilon_j & j = 1, \dots, m & \quad \epsilon_j \sim f_2 \in \mathbb{R}^2, \end{aligned}$$

donde

$$\beta_0 = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix}, \quad B = \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix} \quad \text{y} \quad \Psi(\gamma_0 + y_j^T \gamma) = \begin{pmatrix} \Psi(\gamma_{10} + y_j^T \gamma_1) \\ \Psi(\gamma_{20} + y_j^T \gamma_2) \end{pmatrix}.$$

Análogamente a como se procedió en la sección 4.2.2, se llega a que la función de densidad de cada punto toma la expresión

$$\begin{aligned} f(x_i) &= f_1(x_i - \mu_{\xi_{1i}}) & i = 1, \dots, n \\ f(y_j) &= f_2(\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu_{\xi_{2j}}) |J_j| \\ &= f_2(\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu_{\xi_{2j}}) |\gamma| \prod_{l=1}^2 \beta_l \frac{\Psi(\gamma_{l0} + y_j^T \gamma_l)}{1 + \exp(\gamma_{l0} + y_j^T \gamma_l)} \\ j &= 1, \dots, m, \end{aligned} \tag{4.24}$$

donde

$$\gamma = \begin{pmatrix} \gamma_1^T \\ \gamma_2^T \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \tag{4.25}$$

y

$$\begin{aligned} J_j &= \frac{\partial}{\partial y_j} (\beta_0 + B\Psi(\gamma_0 + y_j^T \gamma) - \mu_{\xi_{2j}}) = B \frac{\partial}{\partial y_j} (\Psi(\gamma_0 + y_j^T \gamma)) \\ &= B \begin{pmatrix} \frac{\Psi(\gamma_{10} + y_j^T \gamma_1)}{1 + \exp(\gamma_{10} + y_j^T \gamma_1)} & 0 \\ 0 & \frac{\Psi(\gamma_{20} + y_j^T \gamma_2)}{1 + \exp(\gamma_{20} + y_j^T \gamma_2)} \end{pmatrix} \gamma. \end{aligned}$$

La verosimilitud de los datos se calcula de la misma forma que en 4.2.2, y a partir de (4.24). Se va diferenciando si los puntos están o no emparejados para calcular la aportación de cada subgrupo de puntos a la verosimilitud total. Así, se calculan las expresiones aproximadas de estas aportaciones, considerando la región $V \subset \mathbb{R}^d$ suficientemente grande con respecto al soporte de f_1 y f_2 .

Asumiendo que los errores siguen una distribución normal y operando como en la sección 4.2.2, la expresión de la verosimilitud de los datos tiene la expresión,

$$\begin{aligned}
 p(\mathbf{x}, \mathbf{y} \mid \mathbf{M}, \beta_0, B, \gamma_0, \gamma, \sigma^2) &\approx |\gamma|^m \prod_{l=1}^2 \beta_l^m \left(\frac{1}{\sigma\sqrt{2}} \right)^{2L} \frac{1}{(\sqrt{2\pi})^L} \\
 &\prod_{\{j: M_{ij}=0 \forall i\}} \prod_{k=1}^2 \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp \{ \gamma_{k0} + y_j^T \gamma_k \})} \cdot \\
 &\prod_{\{(i,j): M_{ij}=1\}} \left[\exp \left\{ -\frac{1}{4\sigma^2} \sum_{l=1}^2 (x_{il} - \beta_{l0} - \beta_l \Psi(\gamma_{l0} + y_j^T \gamma_l))^2 \right\} \cdot \right. \\
 &\left. \prod_{k=1}^2 \frac{\Psi(\gamma_{k0} + y_j^T \gamma_k)}{(1 + \exp \{ \gamma_{k0} + y_j^T \gamma_k \})} \right], \tag{4.26}
 \end{aligned}$$

donde L es el número de emparejamientos entre ambas configuraciones y M_{ij} representa la matriz de emparejamientos definida en (4.3).

4.3.1. Inferencia bayesiana en el caso de configuraciones etiquetadas

Para realizar inferencias sobre los parámetros, nos centramos en el caso en el que las dos configuraciones tienen el mismo número de puntos (n) y que están emparejados, es decir, $L = n$ y, por tanto, se trata de configuraciones etiquetadas.

Sin pérdida de generalidad, supongamos que $\forall i = 1, \dots, n$ los puntos x_i e y_i forman los n emparejamientos, es decir, $\xi_{1i} = \xi_{2i}$. Así,

$$M_{jk} = \begin{cases} 1 & j = k \\ 0 & \text{resto} \end{cases}.$$

A partir de (4.26), prescindiendo de las constantes, la verosimilitud de los datos en el caso de $M = 1$, con errores normales y configuraciones etiquetadas, tiene la expresión,

$$\begin{aligned}
 p(\mathbf{x}, \mathbf{y} \mid \beta_0, B, \gamma_0, \gamma, \sigma^2) &\approx |\gamma|^n \beta_1^n \beta_2^n \left(\frac{1}{\sigma^2} \right)^n \\
 &\exp \left\{ -\frac{1}{4\sigma^2} \sum_{i=1}^n \left[(x_{i1} - \beta_{10} - \beta_1 \Psi(\gamma_{10} + y_i^T \gamma_1))^2 \right. \right. \\
 &\quad \left. \left. + (x_{i2} - \beta_{20} - \beta_2 \Psi(\gamma_{20} + y_i^T \gamma_2))^2 \right] \right\} \\
 &\prod_{i=1}^n \frac{\Psi(\gamma_{10} + y_i^T \gamma_1)}{(1 + \exp \{ \gamma_{10} + y_i^T \gamma_1 \})} \frac{\Psi(\gamma_{20} + y_i^T \gamma_2)}{(1 + \exp \{ \gamma_{20} + y_i^T \gamma_2 \})}. \tag{4.27}
 \end{aligned}$$

A partir de esta verosimilitud y asumiendo distintas distribuciones a priori, se procede al cálculo de las distribuciones a posteriori de los parámetros, cuyas demostraciones no se presentan por ser análogas a las realizadas en la sección 4.2.3.

Distribución a posteriori de σ^2

Asumimos como distribución a priori de la varianza σ^2 una gamma inversa $\sigma^{-2} \sim G(\alpha, \beta)$, $\alpha, \beta > 0$. Entonces utilizando (4.27), e incluyendo en la constante de proporcionalidad todo aquello que no depende de σ^2 ,

$$\begin{aligned}
 p(\sigma^2 \mid \mathbf{x}, \mathbf{y}, \beta_0, B, \gamma_0, \gamma) &\propto (\sigma^2)^{-(\alpha+1)-n} \\
 &\exp \left\{ -\frac{1}{\sigma^2} \left[\beta + \frac{1}{4} \left(\sum_{i=1}^n (x_{i1} - \beta_{10} - \beta_1 \Psi(\gamma_{10} + y_i^T \gamma_1))^2 \right. \right. \right. \\
 &\quad \left. \left. + \sum_{i=1}^n (x_{i2} - \beta_{20} - \beta_2 \Psi(\gamma_{20} + y_i^T \gamma_2))^2 \right) \right] \right\},
 \end{aligned}$$

por lo que se llega a que la distribución a posteriori es $\sigma^{-2} \sim G(\alpha + n, \beta + A)$, donde

$$\begin{aligned}
 A &= \frac{1}{4} \left\{ \sum_{i=1}^n (x_{i1} - \beta_{10} - \beta_1 \Psi(\gamma_{10} + y_i^T \gamma_1))^2 \right. \\
 &\quad \left. + \sum_{i=1}^n (x_{i2} - \beta_{20} - \beta_2 \Psi(\gamma_{20} + y_i^T \gamma_2))^2 \right\}.
 \end{aligned}$$

Distribución a posteriori de β_{10} y β_{20}

Asumiendo que la distribución a priori para $\beta_{r0} \sim N(\mu_{\beta_{r0}}, \sigma_{\beta_{r0}}^2)$, $r = 1, 2$, se obtiene que la distribución a posteriori de β_{r0} es

$$\beta_{r0} \mid \mathbf{x}, \mathbf{y}, \beta_{0(r)}, B, \gamma_0, \boldsymbol{\gamma}, \sigma^2 \sim N\left(\frac{D_{r0}}{2C_{r0}}, C_{r0}^{-1}\right),$$

donde

$$\begin{aligned} C_{r0} &= \frac{1}{\sigma_{\beta_{r0}}^2} + \frac{n}{2\sigma^2}, \\ D_{r0} &= \frac{2\mu_{\beta_{r0}}}{\sigma_{\beta_{r0}}^2} + \frac{\sum_{i=1}^n R_{ir0}}{\sigma^2}, \end{aligned}$$

siendo, en este caso,

$$R_{ir0} = x_{ir} - \beta_r \Psi(\gamma_{r0} + y_i^T \gamma_r) \quad i = 1, \dots, n,$$

expresión independiente de β_{r0} y denotando por $\beta_{0(r)}$ el vector β_0 sin la coordenada β_{r0} .

Distribución a posteriori de β_1 y β_2

Asumiendo que la distribución a priori para $\beta_r \sim N(\mu_{\beta_r}, \sigma_{\beta_r}^2)$, $r = 1, 2$, se obtiene que la distribución a posteriori de β_r tiene la expresión

$$p(\beta_r \mid \mathbf{x}, \mathbf{y}, \beta_0, B_{(r)}, \gamma_0, \boldsymbol{\gamma}, \sigma^2) \propto \beta_r^n \exp\left\{-\frac{1}{2C_r^{-1}}(\beta_r - \frac{D_r}{2C_r})^2\right\}$$

donde

$$\begin{aligned} C_r &= \frac{1}{\sigma_{\beta_r}^2} + \frac{\sum_{i=1}^n \Psi^2(\gamma_{r0} + y_i^T \gamma_r)}{2\sigma^2} \\ D_r &= \frac{\mu_{\beta_r}}{\sigma_{\beta_r}^2} + \frac{\sum_{i=1}^n R_{ir}}{2\sigma^2} \end{aligned}$$

donde

$$R_{ir} = (x_{ir} - \beta_{r0})\Psi(\gamma_{r0} + y_i^T \gamma_r) \quad i = 1, \dots, n,$$

y $B_{(r)}$ denota la matriz B sin β_r .

Distribución a posteriori de γ_{10} y γ_{20}

Para los parámetros γ_{r0} $r = 1, 2$, también asumimos distribuciones normales a priori, es decir, $\gamma_{r0} \sim N(\mu_{\gamma_{r0}}, \sigma_{\gamma_{r0}}^2)$.

De forma análoga a la situación de la sección 4.2.3, se llega a que la expresión de la distribución a posteriori de γ_{r0} es

$$\begin{aligned} p(\gamma_{r0} \mid \mathbf{x}, \mathbf{y}, \beta_0, B, \gamma_{(r)0}, \boldsymbol{\gamma}, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma_{\gamma_{r0}}^2} (\gamma_{r0}^2 - 2\mu_{\gamma_{r0}} \gamma_{r0}) \right. \\ \left. - \frac{1}{4\sigma^2} \beta_r^2 \sum_{i=1}^n \Psi^2(\gamma_{r0} + y_i^T \gamma_r) + \frac{1}{2\sigma^2} \beta_r \sum_{i=1}^n S_{ir0} \Psi(\gamma_{r0} + y_i^T \gamma_r) \right. \\ \left. + n\gamma_{r0} - 2 \sum_{i=1}^n \ln(1 + \exp \{ \gamma_{r0} + y_i^T \gamma_r \}) \right\}, \end{aligned}$$

donde para todo $i = 1, \dots, n$ y $r = 1, 2$

$$S_{ir0} = x_{ir} - \beta_{r0}, \tag{4.28}$$

y $\gamma_{(r)0}$ representa γ_{10} , si $r = 2$, y γ_{20} , si $r = 1$.

Distribución a posteriori de γ_{rs} , $r = 1, 2$ y $s = 1, 2$

Para los parámetros gamma, también se van a suponer distribuciones a priori normales, es decir $\gamma_{rs} \sim N(\mu_{\gamma_{rs}}, \sigma_{\gamma_{rs}}^2)$. Recordamos que corresponden a las coordenadas de dos vectores bidimensionales y son los coeficientes de la combinación lineal de las funciones logísticas.

Basándose en la verosimilitud (4.27) la expresión de la distribución a posteriori de γ_{rs} es

$$p(\gamma_{rs} \mid \mathbf{x}, \mathbf{y}, \beta_0, B, \gamma_0, \boldsymbol{\gamma}_{(rs)}, \sigma^2) \propto |\gamma_{11}\gamma_{22} - \gamma_{21}\gamma_{12}|^n \exp \left\{ -\frac{1}{4\sigma^2} \beta_r^2 \sum_{i=1}^n \Psi^2(\gamma_{r0} + y_i^T \gamma_r) \right. \\ \left. + \frac{1}{2\sigma^2} \beta_r \sum_{i=1}^n S_{ir0} \Psi(\gamma_{r0} + y_i^T \gamma_r) + \gamma_{rs} \sum_{i=1}^n y_{is} \right. \\ \left. - 2 \sum_{i=1}^n L n(1 + \exp \{ \gamma_{r0} + y_i^T \gamma_r \}) - \frac{1}{2\sigma_{\gamma_{rs}}^2} (\gamma_{rs}^2 - 2\mu_{\gamma_{rs}} \gamma_{rs}) \right\},$$

donde S_{ir0} viene definida en (4.28) y $\boldsymbol{\gamma}_{(rs)}$ representa a la matriz $\boldsymbol{\gamma}$ definida en (4.25) salvo el parámetro γ_{rs} .

4.4. Emparejamiento de configuraciones etiquetadas bajo modelo de redes neuronales con una configuración fija

En esta sección y siguiendo parcialmente la idea de Dryden, et al. (2007), vamos a considerar un modelo en el que, a diferencia del desarrollado en la sección 4.2, una de las dos configuraciones es considerada de referencia, (por lo tanto fija), siendo la otra una transformación no lineal de la primera en la que se le incluye también una parte lineal, más un error aleatorio. Esto permite que la transformación lineal entre configuraciones esté incluida en el modelo como un caso particular.

En la figura 4.2 se representa la situación de las configuraciones en este modelo.

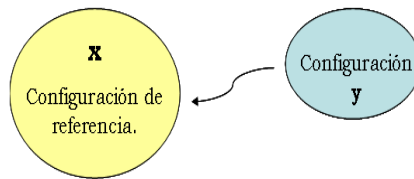


Figura 4.2 Esquema del modelo con una configuración fija

Nos centraremos en el caso de dos configuraciones etiquetadas en \mathbb{R}^2 con emparejamientos conocidos, por lo que el problema se centra en la búsqueda de la red neuronal que mejor relaciona ambas configuraciones.

En primer lugar, se plantea el modelo no lineal de redes neuronales. Después se calcula la función de verosimilitud y se obtienen las expresiones de las distribuciones a posteriori de los parámetros.

4.4.1. Modelización del problema

Inicialmente se consideran las configuraciones de n puntos en \mathbb{R}^d ,

$$\mathbf{x} = \{x_i \mid i = 1, \dots, n\} \text{ e } \mathbf{y} = \{y_i \mid i = 1, \dots, n\}.$$

Se supone que para todo $i = 1, \dots, n$, los puntos (x_i, y_i) están emparejados y se considera la configuración \mathbf{x} fija e \mathbf{y} aleatoria cuya distribución depende de la primera mediante el modelo no lineal,

$$y_{ij} = \beta_{j0} + \lambda_j^T x_i + \sum_{k=1}^M \beta_{jk} \Psi(\gamma_{k0} + x_i^T \gamma_k) + \varepsilon_{ij}, \quad (4.29)$$

donde para todo $i = 1, \dots, n$, $j = 1, \dots, d$ y $k = 1, \dots, M$, los parámetros son $\beta_{j0} \in \mathbb{R}$, $\lambda_j \in \mathbb{R}^d$, $\beta_{jk} \in \mathbb{R}$, $\gamma_{k0} \in \mathbb{R}$ y $\gamma_k \in \mathbb{R}^d$, los errores ε_{ij} siguen distribución f_i y la función $\Psi(z)$ representa la función logística.

Con este modelo se expresa cada coordenada $j = 1, \dots, d$ del punto i -ésimo de la configuración \mathbf{y} , y_{ij} , como la suma de un término lineal de x_i , (con coeficientes el vector λ_j), y de una combinación lineal de M términos, que son funciones logísticas de combinaciones lineales de las coordenadas del punto x_i , más un error aleatorio. Los coeficientes que definen esta combinación lineal de funciones logísticas se recogen en el vector $\beta_j^T = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jM})$.

Se asume en este modelo que el número de nodos M es conocido, aunque posteriormente, en la aplicación con datos reales, se llevará a cabo un análisis de los modelos con distintos valores de M basado en criterios de información.

El modelo (4.29) se puede expresar matricialmente,

$$y_i = \beta_{00} + \Lambda x_i^T + B\Psi(\gamma_0 + x_i^T \gamma) + \varepsilon_i,$$

donde $i = 1, \dots, n$, $\varepsilon_i \sim f_i \in \mathbb{R}^d$ independientes, y B y $\Psi(\gamma_0 + x_i^T \gamma)$ vienen definidas en (4.2) y

$$\beta_{00} = \begin{pmatrix} \beta_{10} \\ \vdots \\ \beta_{d0} \end{pmatrix}_{d \times 1} \quad \Lambda = \begin{pmatrix} \lambda_1^T \\ \vdots \\ \lambda_d^T \end{pmatrix}_{d \times d} = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1d} \\ \vdots & & \vdots \\ \lambda_{d1} & \cdots & \lambda_{dd} \end{pmatrix}.$$

Asumiendo normalidad de los errores, el modelo se puede expresar como,

$$y_i = \beta_{00} + \Lambda x_i^T + B\Psi(\gamma_0 + x_i^T \gamma) + \varepsilon_i \quad i = 1, \dots, n \quad \varepsilon_i \sim \mathbf{N}_d(0, \sigma^2 \mathbf{I}_d). \quad (4.30)$$

4.4.2. Construcción de la verosimilitud de los datos

Basándonos en la normalidad de los errores se tiene que para todo $i = 1, \dots, n$,

$$f(y_i) = f_i(y_i - \beta_{00} \Lambda x_i^T + B\Psi(\gamma_0 + x_i^T \gamma)) = \frac{1}{(\sqrt{2\pi})^d (\sigma^2)^{d/2}} \exp\left\{-\frac{1}{2\sigma^2} (y_i - \beta_{00} \Lambda x_i^T + B\Psi(\gamma_0 + x_i^T \gamma))^T (y_i - \beta_{00} \Lambda x_i^T + B\Psi(\gamma_0 + x_i^T \gamma))\right\},$$

de modo que la verosimilitud es,

$$p(\mathbf{y} \mid \beta_{00}, B, \Lambda, \gamma_0, \gamma, \sigma^2) \propto \frac{1}{(\sigma^2)^{\frac{nd}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \sum_{j=1}^d (y_{ij} - \beta_{j0} - \lambda_j^T x_i - \sum_{k=1}^M \beta_{jk} \Psi(\gamma_{k0} + x_i^T \gamma_k))^2\right\}.$$

En el caso particular donde $d = 2$, la expresión de la función de verosimilitud toma la forma:

$$\begin{aligned} p(\mathbf{y} \mid \beta_{00}, B, \Lambda, \gamma_0, \gamma, \sigma^2) &\propto \frac{1}{\sigma^{2n}} \cdot \\ &\exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (y_{i1} - \beta_{10} - \lambda_1^T x_i - \sum_{k=1}^M \beta_{1k} \Psi(\gamma_{k0} + x_i^T \gamma_k))^2 \right. \right. \\ &\quad \left. \left. + \sum_{i=1}^n (y_{i2} - \beta_{20} - \lambda_2^T x_i - \sum_{k=1}^M \beta_{2k} \Psi(\gamma_{k0} + x_i^T \gamma_k))^2 \right] \right\}. \end{aligned} \quad (4.31)$$

4.4.3. Expresiones de las distribuciones a posteriori de los parámetros

Distribución a posteriori de σ^2

Asumimos que la distribución a priori de σ^2 es una gamma inversa, es decir, $\sigma^{-2} \sim G(\alpha, \beta)$. A partir de (4.31) la distribución a posteriori de σ^2 será de la forma,

$$\begin{aligned} p(\sigma^2 \mid \mathbf{y}, \beta_{00}, B, \Lambda, \gamma_0, \boldsymbol{\gamma}) &= p(\sigma^2 \mid \mathbf{y}) \propto (\sigma^2)^{-(\alpha+1)} \exp \left\{ -\frac{\beta}{\sigma^2} \right\} \cdot \\ &\left(\frac{1}{\sigma^2} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (y_{i1} - \beta_{10} - \lambda_1^T x_i - \sum_{k=1}^M \beta_{1k} \Psi(\gamma_{k0} + x_i^T \gamma_k))^2 \right. \right. \\ &\quad \left. \left. + \sum_{i=1}^n (y_{i2} - \beta_{20} - \lambda_2^T x_i - \sum_{k=1}^M \beta_{2k} \Psi(\gamma_{k0} + x_i^T \gamma_k))^2 \right] \right\} \\ &= (\sigma^2)^{-(\alpha+1+n)} \exp \left\{ -\frac{1}{\sigma^2} (\beta + A) \right\}, \end{aligned}$$

donde

$$\begin{aligned} A &= \frac{1}{2} \left[\sum_{i=1}^n (y_{i1} - \beta_{10} - \lambda_1^T x_i - \sum_{k=1}^M \beta_{1k} \Psi(\gamma_{k0} + x_i^T \gamma_k))^2 \right. \\ &\quad \left. + \sum_{i=1}^n (y_{i2} - \beta_{20} - \lambda_2^T x_i - \sum_{k=1}^M \beta_{2k} \Psi(\gamma_{k0} + x_i^T \gamma_k))^2 \right]. \end{aligned}$$

Por lo tanto, la distribución a posteriori de σ^2 es una gamma inversa de parámetros,

$$\sigma^{-2} \mid \mathbf{y}, \beta_{00}, B, \Lambda, \gamma_0, \boldsymbol{\gamma} \sim G(\alpha + n, \beta + A)$$

Distribución a posteriori de β_{r0} $r = 1, 2$.

Asumimos que la distribución a priori para $\beta_{r0} \sim N(\mu_{\beta_{r0}}, \sigma_{\beta_{r0}}^2)$. Entonces, basándonos en la función de verosimilitud (4.31) la distribución a posteriori tendrá la

forma,

$$p(\beta_{r0} \mid \mathbf{y}, \beta_{00(r)}, B, \Lambda, \gamma_0, \boldsymbol{\gamma}, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma_{\beta_{r0}}^2} (\beta_{r0}^2 - 2\mu_{\beta_{r0}}\beta_{r0}) \right\} \cdot \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_{ir} - \beta_{r0} - \lambda_1^T x_i - \sum_{k=1}^M \beta_{rk} \Psi(\gamma_{k0} + x_i^T \gamma_k))^2 \right\}, \quad (4.32)$$

donde $\beta_{00(r)}$ denota β_{10} si $r = 2$ y β_{20} si $r = 1$.

Identificando en la expresión lo que es independiente de β_{r0} , para todo $i = 1, \dots, n$ se define

$$R_{ir0} = y_{ir} - \lambda_r^T x_i - \sum_{k=1}^M \beta_{rk} \Psi(\gamma_{k0} + x_i^T \gamma_k). \quad (4.33)$$

Operando de la misma manera que en la sección 4.2.3 se llega a que la distribución a posteriori de β_{r0} es normal

$$\beta_{r0} \mid \mathbf{y}, \beta_{00(r)}, B, \Lambda, \gamma_0, \boldsymbol{\gamma}, \sigma^2 \sim N\left(\frac{D_{r0}}{2C_{r0}}, C_{r0}^{-1}\right),$$

donde

$$C_{r0} = \frac{1}{\sigma_{\beta_{r0}}^2} + \frac{n}{\sigma^2} \quad \text{y}$$

$$D_{r0} = \frac{2\mu_{\beta_{r0}}}{\sigma_{\beta_{r0}}^2} + \frac{2 \sum_{i=1}^n R_{ir0}}{\sigma^2},$$

siendo para todo $i = 1, \dots, n$,

$$R_{ir0} = y_{ir} - \lambda_r^T x_i - \sum_{k=1}^M \beta_{rk} \Psi(\gamma_{k0} + x_i^T \gamma_k).$$

Distribución a posteriori de λ_{rs} $r, s = 1, 2$.

Asumimos que la distribución a priori de λ_{rs} es $\lambda_{rs} \sim N(\mu_{\lambda_{rs}}, \sigma_{\lambda_{rs}}^2)$. Entonces, para todo $r = 1, 2$ y $s = 1, 2$, y denotando por $\Lambda_{(rs)}$ la matriz Λ sin λ_{rs} ,

$$p(\lambda_{rs} \mid \mathbf{y}, \beta_{00}, B, \Lambda_{(rs)}, \gamma_0, \boldsymbol{\gamma}, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma_{\lambda_{rs}}^2} (\lambda_{rs}^2 - 2\mu_{\lambda_{rs}} \lambda_{rs}) \right\} \cdot \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_{ir} - \beta_{r0} - \lambda_{r1}x_{i1} - \lambda_{r2}x_{i2} - \sum_{k=1}^M \beta_{rk} \Psi(\gamma_{k0} + x_i^T \boldsymbol{\gamma}_k))^2 \right\}. \quad (4.34)$$

Si denominamos para todo $i = 1, \dots, n$

$$M_{ir1} = y_{ir} - \beta_{r0} - \lambda_{r2}x_{i2} - \sum_{k=1}^M \beta_{rk} \Psi(\gamma_{k0} + x_i^T \boldsymbol{\gamma}_k),$$

y

$$M_{ir2} = y_{ir} - \beta_{r0} - \lambda_{r1}x_{i1} - \sum_{k=1}^M \beta_{rk} \Psi(\gamma_{k0} + x_i^T \boldsymbol{\gamma}_k),$$

expresiones independientes de λ_{r1} y λ_{r2} , respectivamente, (4.34) se puede escribir,

$$p(\lambda_{rs} \mid \mathbf{y}, \beta_{00}, B, \Lambda_{(rs)}, \gamma_0, \boldsymbol{\gamma}, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma_{\lambda_{rs}}^2} (\lambda_{rs}^2 - 2\mu_{\lambda_{rs}} \lambda_{rs}) \right\} \cdot \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (M_{irs} - \lambda_{rs}x_{is})^2 \right\}.$$

Desarrollando el cuadrado y simplificando se obtiene

$$p(\lambda_{rs} \mid \mathbf{y}, \beta_{00}, B, \Lambda_{(rs)}, \gamma_0, \boldsymbol{\gamma}, \sigma^2) \propto \exp \left\{ -\frac{1}{2} \left[\lambda_{rs}^2 \left(\frac{\sum_{i=1}^n x_{is}^2}{\sigma^2} + \frac{1}{\sigma_{\lambda_{rs}}^2} \right) - 2\lambda_{rs} \left(\frac{\sum_{i=1}^n M_{irs}x_{is}}{\sigma^2} + \frac{\mu_{\lambda_{rs}}}{\sigma_{\lambda_{rs}}^2} \right) \right] \right\}.$$

Llamando

$$C_{rs} = \frac{\sum_{i=1}^n x_{is}^2}{\sigma^2} + \frac{1}{\sigma_{\lambda_{rs}}^2} \quad y \quad (4.35)$$

$$D_{rs} = \frac{\sum_{i=1}^n M_{irs}x_{is}}{\sigma^2} + \frac{\mu_{\lambda_{rs}}}{\sigma_{\lambda_{rs}}^2}, \quad (4.36)$$

y completando cuadrados, se obtiene que la expresión de la distribución a posteriori de λ_{rs} es

$$p(\lambda_{rs} \mid \mathbf{y}, \beta_{00}, B, \Lambda_{(rs)}, \gamma_0, \boldsymbol{\gamma}, \sigma^2) \propto \exp \left\{ -\frac{1}{2C_{rs}^{-1}} \left(\lambda_{rs} - \frac{D_{rs}}{C_{rs}} \right)^2 \right\},$$

que corresponde a una distribución normal,

$$\lambda_{rs} \mid \mathbf{y}, \beta_{00}, B, \Lambda_{(rs)}, \gamma_0, \boldsymbol{\gamma}, \sigma^2 \sim N\left(\frac{D_{rs}}{C_{rs}}, C_{rs}^{-1}\right),$$

donde C_{rs} y D_{rs} vienen definidos en (4.35) y (4.36), respectivamente.

Distribución a posteriori de β_{rs} $r = 1, 2$, $s = 1, \dots, M$

Se asume para β_{rs} una distribución a priori normal $\beta_{rs} \sim N(\mu_{\beta_{rs}}, \sigma_{\beta_{rs}}^2)$, donde $r = 1, 2$ y $s = 1, \dots, M$. Entonces, dada la función de verosimilitud (4.31) la distribución a posteriori tendrá la forma,

$$\begin{aligned} p(\beta_{rs} \mid \mathbf{y}, \beta_{00}, B_{(rs)}, \Lambda, \gamma_0, \boldsymbol{\gamma}, \sigma^2) &\propto \exp \left\{ -\frac{1}{2\sigma_{\beta_{rs}}^2} (\beta_{rs}^2 - 2\mu_{\beta_{rs}} \beta_{rs}) \right\} \cdot \\ &\exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_{ir} - \beta_{r0} - \lambda_r^T x_i - \right. \\ &\quad \left. - \sum_{\substack{k=1 \\ k \neq s}}^M \beta_{rk} \Psi(\gamma_{k0} + x_i^T \gamma_k) - \beta_{rs} \Psi(\gamma_{s0} + x_i^T \gamma_s))^2 \right\}. \end{aligned}$$

Si identificamos como R_{irs} el sumando i de la exponencial

$$R_{irs} = y_{ir} - \beta_{r0} - \lambda_r^T x_i - \sum_{\substack{k=1 \\ k \neq s}}^M \beta_{rk} \Psi(\gamma_{k0} + x_i^T \gamma_k), \quad (4.37)$$

y operando de la misma forma que se hacía en la sección 4.2.3, se llega a que la distribución a posteriori de β_{rs} es normal,

$$\beta_{rs} \mid \mathbf{y}, \beta_{00}, B_{(rs)}, \Lambda, \gamma_0, \boldsymbol{\gamma}, \sigma^2 \sim N\left(\frac{D_{rs}}{C_{rs}}, C_{rs}^{-1}\right),$$

donde, en este caso,

$$\begin{aligned}
 C_{rs} &= \frac{\sum_{i=1}^n \Psi^2(\gamma_{s0} + x_i^T \gamma_s)}{\sigma^2} + \frac{1}{\sigma_{\beta_{rs}}^2}, \\
 D_{rs} &= \frac{\sum_{i=1}^n R_{irs} \Psi(\gamma_{s0} + x_i^T \gamma_s)}{\sigma^2} + \frac{\mu_{\beta_{rs}}}{\sigma_{\beta_{rs}}^2} \quad \text{y} \\
 R_{irs} &= y_{ir} - \beta_{r0} - \lambda_r^T x_i - \sum_{\substack{k=1 \\ k \neq s}}^M \beta_{rk} \Psi(\gamma_{k0} + x_i^T \gamma_k).
 \end{aligned}$$

Distribución a posteriori de γ_{rs} $r = 1, \dots, M$ y $s = 0, 1, 2$

Se asume para γ_{rs} distribuciones a priori normales $\gamma_{rs} \sim N(\mu_{\gamma_{rs}}, \sigma_{\gamma_{rs}}^2)$. Procediendo al cálculo de la distribución a posteriori de forma análoga a la sección 4.2.3 se llega a que su expresión es de la forma

$$\begin{aligned}
 p(\gamma_{rs} \mid \mathbf{y}, \beta_{00}, B, \Lambda, \gamma_0, \boldsymbol{\gamma}_{(rs)}, \sigma^2) &\propto \exp\left\{-\frac{1}{2\sigma^2} \sum_{s=1}^2 \sum_{i=1}^n \beta_{sr}^2 \Psi^2(\gamma_{r0} + x_i^T \gamma_r)\right. \\
 &\quad \left. + \frac{1}{\sigma^2} \sum_{s=1}^2 \sum_{i=1}^n \beta_{sr} R_{irs} \Psi(\gamma_{r0} + x_i^T \gamma_r) - \frac{1}{2\sigma_{\gamma_{rs}}^2} (\gamma_{rs}^2 - 2\mu_{\gamma_{rs}} \gamma_{rs})\right\}, \quad (4.38)
 \end{aligned}$$

donde R_{irs} está definida en (4.37) y $\boldsymbol{\gamma}_{(rs)}$ denota todos los elementos $\boldsymbol{\gamma}$ excepto γ_{rs} .

Por último, la expresión (4.38) debe considerarse, en el caso $s = 0$, como función de γ_{r0} , en el caso $s = 1$, como función de γ_{r1} , y en el caso de $s = 2$, como función de γ_{r2} .

A partir de las distribuciones a posteriori de los parámetros, podremos aplicar el muestreador de Gibbs para simular una muestra de la distribución a posteriori conjunta de los mismos. Todas las distribuciones a posteriori son normales o gamma inversa, excepto las distribuciones de los parámetros gamma. En este caso, se utiliza un paso Metropolis-Hasting tomando como distribución propuesta la distribución normal con varianza adaptativa.

4.5. Una aplicación en Bioinformática

Una vez presentados los resultados teóricos obtenidos con dos planteamientos diferentes para el estudio de emparejamientos entre configuraciones con transformaciones no lineales, en esta sección consideramos una aplicación de ambos modelos, al análisis de las expresiones de una serie de genes recogidos en microarrays. En concreto, nos centraremos en la estimación de los parámetros del modelo, en ambos planteamientos, pues partiremos de dos configuraciones etiquetadas con puntos emparejados que representarán los genes en estudio.

Para ello se va a utilizar la base de datos de microarrays ALL (Acute Lymphoblastic Leukemia) de Chiaretti et al. (2004) de pacientes con leucemia linfoblástica aguda, disponible en el paquete ALL de Bioconductor. Se dispone de los valores de las expresiones de un total de 12625 genes en 128 muestras, cada una correspondiente a un paciente. Las puntuaciones de las expresiones de cada gen en cada una de las muestras se generaron usando la tecnología de *Affymetrix Genechip 5.0*.

Siguiendo a Gentleman et al. (2005), consideramos un subconjunto de 79 pacientes con leucemia linfoblástica aguda tipo B, 42 de ellos citogenéticamente normales y los 37 restantes con una cierta alteración genética denominada fusión genética BCR/ABL. En parte de los cromosomas 9 y 22 se producen intercambios, de manera que parte del gen de región de fractura (BCR, Breakpoint Cluster Region) del cromosoma 22 se fusiona con parte del gen ABL del cromosoma 9.

Por otra parte, siguiendo a Gentleman et al. (2005) descartamos aquellos genes que presentan baja variabilidad entre las muestras. En concreto, se filtran genes con una expresión mayor que 100 en al menos el 25 % de las muestras y un rango intercuartílico de al menos 0.5.

Además, se realizan comparaciones múltiples de las expresiones medias de todos los genes, basándose en el criterio *FDR* (*False Discovery Rate*) a un nivel de 0.05 de modo que se fija la proporción esperada de falsos positivos (o falsas significaciones)

entre los genes que son significativamente diferentes (véase Benjamini y Hochberg, 1995). Así, seleccionamos aquellos genes que muestran diferencias significativas entre los dos grupos, quedándonos con un número final de 102 genes.

En cuanto a cuestiones computacionales para el tratamiento de la base de datos ALL y la selección de genes, se utilizaron los paquetes *genefilter* y *multtest* de Bioconductor.

4.5.1. Representación de los genes

De la misma manera que se procedió en la sección 3.7.2 se ha realizado un análisis INDSCAL, utilizando SAS v. 9.1, usando las distancias euclídeas entre genes para cada uno de los dos grupos de pacientes con leucemia. Así, se obtienen las coordenadas de los 102 genes en un espacio de dos dimensiones y los pesos que cada grupo asigna a cada dimensión.

Se obtienen pues, dos configuraciones de 102 puntos, una correspondiente a los pacientes citogenéticamente normales y otra a los que presentan la fusión genética BCR/ABL, donde cada punto identifica a un gen. Los emparejamientos de los puntos en ambas configuraciones son conocidos, correspondiendo al mismo gen.

En las figuras 4.3 y 4.4 se representan, respectivamente, el espacio global de objetos y el espacio de sujetos. Se observa cómo el grupo BCR/ABL da más peso a la dimensión uno que el grupo de los pacientes citogenéticamente normales.

Las figuras 4.5 y 4.6 representan individualmente las configuraciones de los 102 genes en cada grupo. De nuevo, se han identificado dos genes únicamente a efectos ilustrativos.

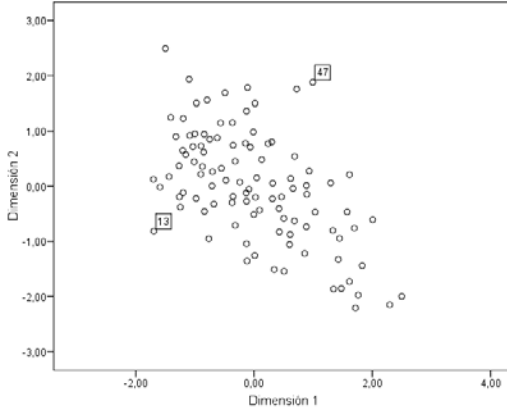


Figura 4.3 Espacio global de objetos

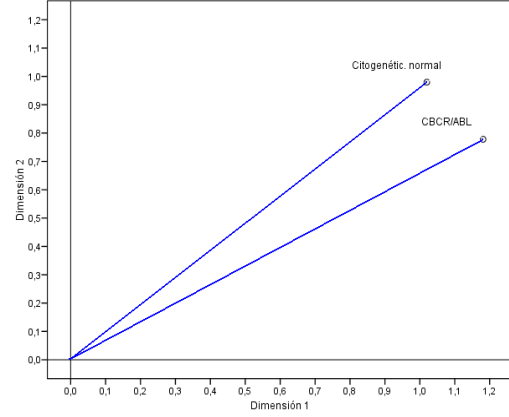


Figura 4.4 Espacio de sujetos

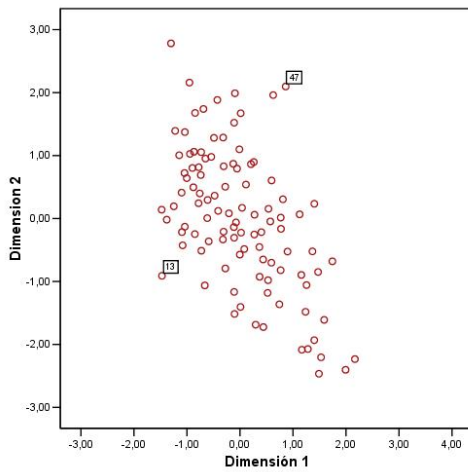


Figura 4.5 Mapa de genes en el grupo
citogenéticamente normal

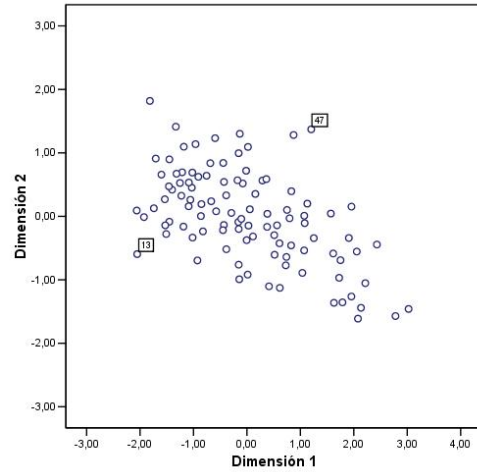


Figura 4.6. Mapa de genes en el grupo
con fusión genética BCR/ABL

4.5.2. Elección del número de nodos del modelo

En cualquiera de los dos modelos de redes neuronales presentados en las secciones 4.2 y 4.4, se puede introducir como un parámetro más del mismo, al número de nodos M . Algunos autores (véanse, e.g. Ríos-Insua, (1998) y Andrieu et al. (2001)), proponen modelos de salto reversible para explorar entre espacios paramétricos de distintas dimensiones. Sin embargo, en modelos complejos como los de redes neuronales, pueden aparecer problemas de identificabilidad lo cual complica la utilización de los

métodos de salto reversible.

En esta aplicación hemos optado por resolver este problema eligiendo, en cada situación, aquel modelo con el que mejor medida de parsimonia tenga, según el criterio DIC (Deviance Information Criterium). Nos hemos basado en una versión alternativa de esta medida propuesta por Richardson (2002) y que Celeux et al. (2006) denotaron por DIC_3 que se adapta bien a la estructura de los modelos de redes neuronales y que viene definida por:

$$\widehat{DIC}_3 = -4E_{\theta|\mathbf{y}}[\log f(\mathbf{y} \mid \theta)] + 2 \log \hat{f}(\mathbf{y}),$$

donde $\hat{f}(\mathbf{y}) = \prod_{i=1}^n \hat{f}(y_i)$, y $\hat{f}(y_i) = E_{\theta|\mathbf{y}}[f(y_i \mid \theta)]$.

Utilizaremos esta medida DIC_3 como criterio de comparación de modelos con distintos valores de M , procediendo a la elección del modelo con menor valor de DIC_3 .

Aplicaremos un paso Metropolis-Hasting para simular una muestra de la distribución a posteriori conjunta de los parámetros. Se ha utilizado el programa *WinBUGS* 1.4.3 para programar el modelo y se ha ejecutado desde R utilizando el paquete *R2WinBUGS*.

A continuación se presentan las inferencias sobre los parámetros de los dos modelos estudiados en las secciones 4.2 y 4.4, particularizados en el caso de las dos configuraciones de genes.

4.5.3. Aplicación del modelo de redes neuronales basado en un Proceso de Poisson

Como aplicación del modelo (4.1), desarrollado en el caso de configuraciones con puntos emparejados y bajo normalidad (sección 4.2.3), tomamos como configuración \mathbf{x} la correspondiente a los genes de los pacientes citogenéticamente normales siendo \mathbf{y} la configuración de los genes de los pacientes con alteración genética.

Se asumen distribuciones a priori de los parámetros $N(0, 1)$, (no se espera unos valores muy elevados de los parámetros al tener estandarizadas las observaciones), excepto para σ^{-2} que se asume una $G(0.1, 0.1)$, y se elige como modelo aquel con el valor del DIC_3 más pequeño.

En la tabla 4.1 se representan los valores de los DIC_3 en los modelos con distintos valores de M . En el caso de $M = 1$, se ha estimado el modelo presentado en la sección 4.3. El menor valor del DIC_3 se obtiene en el caso $M = 2$ nodos con un valor de 76,83.

M	1	2	3	4
DIC_3	2368,43	76,83	436,64	816,54
M	5	6	7	8
DIC_3	1182,409	1561,37	1966,03	2360,15
M	9	10		
DIC_3	2767,181	3193,18		

Tabla 4.1 Tabla de los DIC_3 en función de M

Por lo tanto se elige como modelo para representar la relación entre las dos configuraciones de genes,

$$x_i = \mu_i + e_i$$

$$\left(\begin{array}{c} \beta_{10} + \beta_{11} \frac{1}{1 + \exp\{-(\gamma_{10} + \gamma_1^T y_i)\}} + \beta_{12} \frac{1}{1 + \exp\{-(\gamma_{20} + \gamma_2^T y_i)\}} \\ \beta_{20} + \beta_{21} \frac{1}{1 + \exp\{-(\gamma_{10} + \gamma_1^T y_i)\}} + \beta_{22} \frac{1}{1 + \exp\{-(\gamma_{20} + \gamma_2^T y_i)\}} \end{array} \right) = \mu_i + \varepsilon_i \quad i = 1, \dots, n$$

donde e_i y ε_i siguen distribución normal y son independientes.

La simulación de la muestra de la distribución a posteriori conjunta de los parámetros, se ha llevado a cabo con 50000 iteraciones en total, 25000 iteraciones de calentamiento. Como estimaciones se toman las medias de las distribuciones a posteriori, cuyos valores se recogen en la tabla 4.2 junto a los intervalos de probabilidad al 95 %.

Parámetro	Estimación	Intervalo al 95 %
β_{10}	-1.755	(-2.403 , -1.022)
β_{20}	-1.298	(-2.258 , -0.416)
β_{11}	4.259	(3.519 , 5.081)
β_{12}	-0.159	(-1.605 , 1.063)
β_{21}	-1.425	(-2.769 , -0.125)
β_{22}	4.049	(3.171 , 5.030)
γ_{10}	-0.324	(-0.653 , 0.087)
γ_{20}	-0.148	(-0.469 , 0.181)
γ_{11}	1.079	(0.811 , 1.423)
γ_{12}	0.368	(0.030 , 0.723)
γ_{21}	0.048	(-0.301 , 0.448)
γ_{22}	1.135	(0.8816 , 1.456)
σ^2	0.08	(0.028 , 0.360)

Tabla 4.2 Estimación de los parámetros

Con respecto a la convergencia de las cadenas, en las figuras 4.7 hasta 4.15 se muestran los gráficos de las trazas de las cadenas, observándose la estabilidad de las mismas.

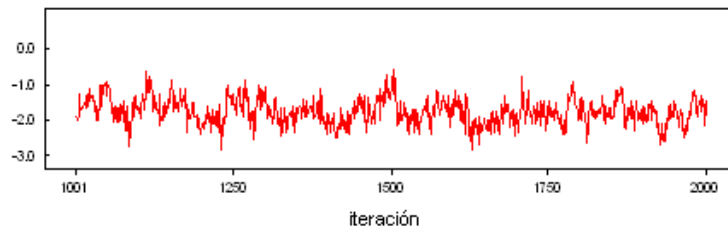


Figura 4.7 Traza de β_{10}

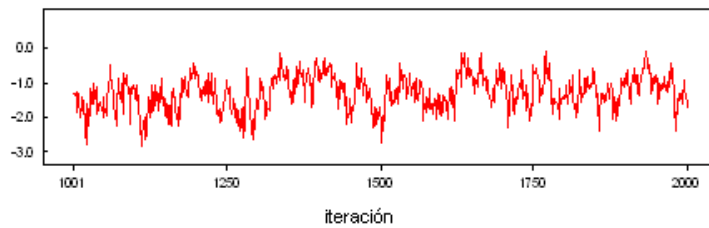


Figura 4.8 Traza de β_{20}

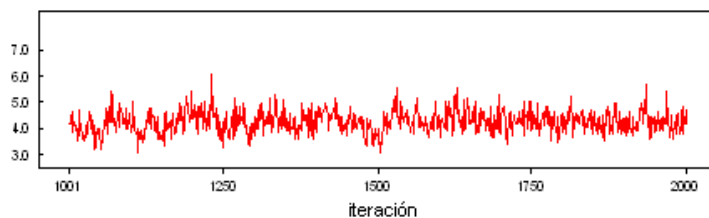


Figura 4.9 Traza de β_{11}

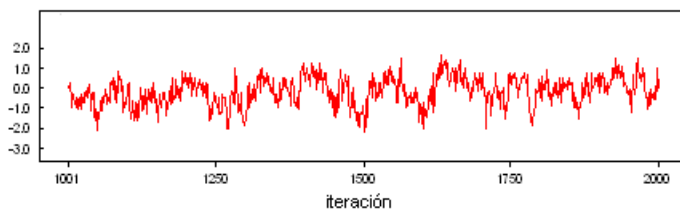


Figura 4.10 Traza de β_{12}

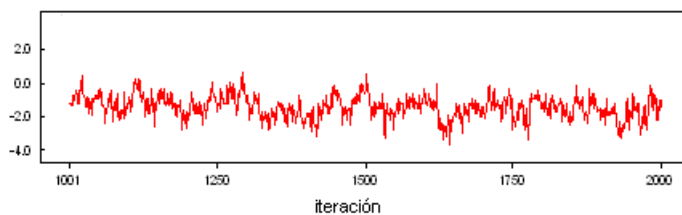


Figura 4.11 Traza de β_{21}

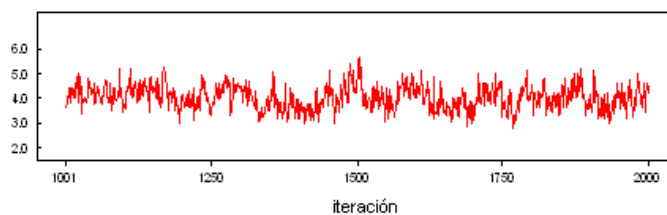


Figura 4.12 Traza de β_{22}

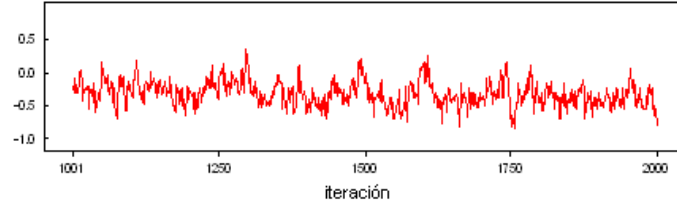


Figura 4.13 Traza de γ_{10}

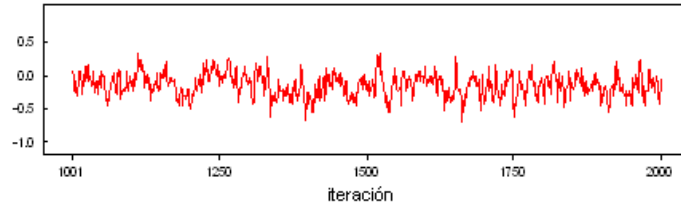


Figura 4.14 Traza de γ_{20}

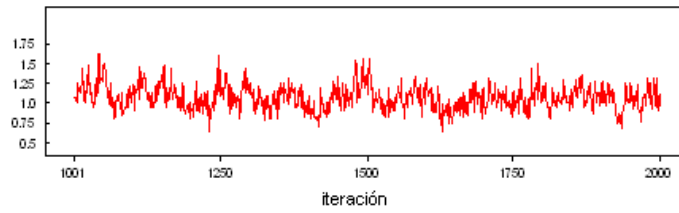


Figura 4.15 Traza de γ_{11}

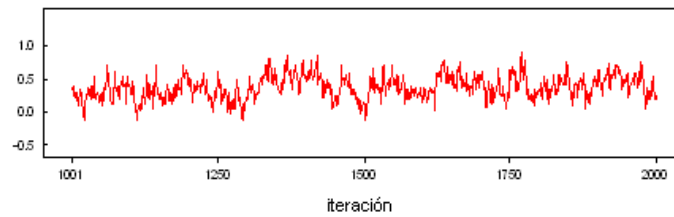


Figura 4.16 Traza de γ_{12}

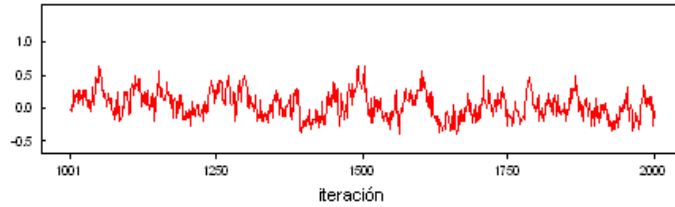


Figura 4.17 Traza de γ_{21}

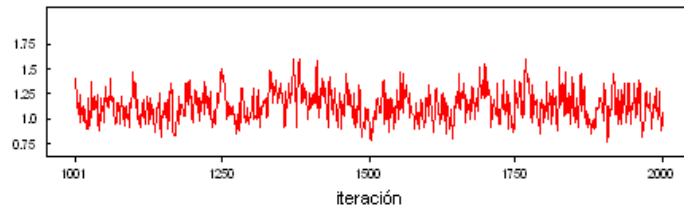


Figura 4.18 Traza de γ_{22}

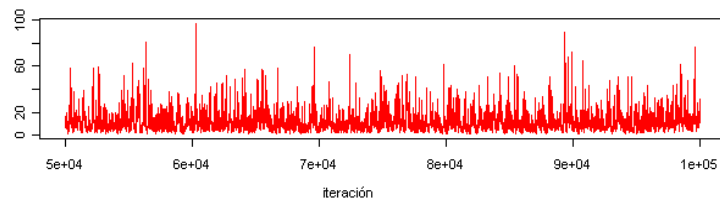


Figura 4.19 Traza de σ^2

Además usamos el método de diagnóstico de Geweke (1992) que se aplica a una sola cadena. Consiste básicamente en comparar el valor medio de los primeros valores de la cadena (por ejemplo, el 10 %), con el valor medio de los últimos (por ejemplo, 50 %). Si los dos tramos de la cadena provienen de la misma distribución estacionaria, el estadístico de Geweke tiene como distribución límite una normal estándar. Un valor no significativo del test implica estabilidad de la cadena.

Los valores de los p -valores asociados a los estadísticos de Geweke de cada parámetro del modelo se muestran en la tabla 4.3, obteniéndose no significación en todos ellos, y por tanto, demostrando la estacionariedad de todas las cadenas.

Parámetro	Z-valor de Geweke	p-valor
β_{10}	1.244	0.215
β_{20}	-1.948	0.052
β_{11}	0.418	0.682
β_{12}	-0.934	0.352
β_{21}	0.405	0.689
β_{22}	-0.042	0.968
γ_{10}	1.250	0.211
γ_{20}	1	0.317
γ_{11}	-0.250	0.802
γ_{12}	1.063	0.289
γ_{21}	-1.21	0.226
γ_{22}	0.490	0.624
σ^2	-0.231	0.818

Tabla 4.2 Estimación de los parámetros

4.5.4. Aplicación del modelo de redes neuronales con una configuración de referencia

En este caso, aplicamos el modelo (4.30) tomando como configuración fija y de referencia la correspondiente a los enfermos de leucemia citogenéticamente normales.

En la figura 4.20 se representan los valores obtenidos del DIC_3 en modelos con distintos valores de M . El menor valor del DIC_3 se obtiene en el caso de $M = 1$ nodo con un valor del $DIC_3 = -1494,56$, por lo que se elegirá el modelo con un único nodo. También se calculó el DIC_3 para el modelo lineal, es decir $M = 0$ nodos, obteniéndose un valor del $DIC_3 = -1024,46$, valor mayor que con $M = 1, \dots, 10$.

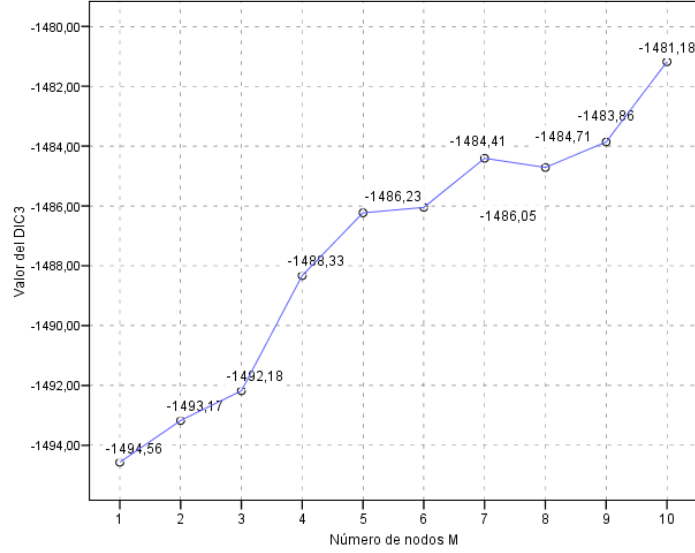


Figura 4.20 Representación de los DIC₃ en función de M

Por lo tanto, el modelo con el que se realizarán las estimaciones de los parámetros será, $\forall i = 1, \dots, n$,

$$y_{i1} = \beta_{10} + \lambda_{11}x_{i1} + \lambda_{12}x_{i2} + \beta_1 \frac{1}{1 + e^{-(\gamma_{10} + \gamma_1 x_{i1} + \gamma_2 x_{i2})}} + \varepsilon_{i1},$$

$$y_{i2} = \beta_{20} + \lambda_{21}x_{i1} + \lambda_{22}x_{i2} + \beta_2 \frac{1}{1 + e^{-(\gamma_{20} + \gamma_1 x_{i1} + \gamma_2 x_{i2})}} + \varepsilon_{i2} \quad \varepsilon_{ij} \sim N(0, \sigma^2).$$

Estimación, validación del modelo y análisis de convergencia

Para comprobar la validez de los resultados se optó por dividir la muestra de los 102 genes en dos submuestras, una para realizar la estimación de los parámetros y otra para realizar la validación. Se seleccionó una muestra aleatoria del 20 % de los puntos (un total de 20 genes) para construir la muestra de validación y se procedió a la estimación de los parámetros del modelo con el 80 % de los genes restantes.

De nuevo, aplicamos el MCMC usando *Winbugs*, simulando tres cadenas con 300000 iteraciones, 150000 de calentamiento, para cada parámetro. Como estimaciones de las predicciones se toman las medianas a posteriori, porque resultó ser una medida más robusta como valor central.

CAPÍTULO 4. EMPAREJAMIENTO DE PARES DE CONFIGURACIONES MEDIANTE TRANSFORMACIONES NO LINEALES

Las distribuciones a priori que se tomaron para los parámetros fueron $N(0, 10)$, excepto para σ^{-2} donde se tomó una $G(10, 10)$, lo que implica que asumimos que, a priori, no se espera gran variabilidad aleatoria entre los genes.

En la tabla 4.4 se muestran los valores de las estimaciones de las coordenadas de los 20 genes utilizados como submuestra de validación. Se observa un buen ajuste, donde las diferencias entre los valores reales y las estimaciones son del orden de centésimas.

Valores reales	(0.6085 , -0.4257)	(1.1290 , 0.1995)
Estimaciones	(0.6142 , -0.4257)	(1.1380 , 0.2034)
Valores reales	(-0.8526 , 0.1925)	(-1.0322 , 0.4513)
Estimaciones	(-0.8519 , 0.1885)	(-1.0300 , 0.4472)
Valores reales	(1.0799 , -0.1076)	(1.9554 , 0.1529)
Estimaciones	(1.0860 , -0.1073)	(1.9640 , 0.1610)
Valores reales	(-0.9272 , -0.6950)	(-0.4438 , 0.8406)
Estimaciones	(-0.9232 , -0.6994)	(-0.4345 , 0.8381)
Valores reales	(0.1124 , -0.3172)	(-1.0589 , 0.2600)
Estimaciones	(0.1144 , -0.3191)	(-1.0590 , 0.2540)
Valores reales	(-1.0176 , -0.3346)	(-0.0058 , -0.3753)
Estimaciones	(-1.0160 , -0.3369)	(-0.0037 , -0.3778)
Valores reales	(-1.7069 , 0.9099)	(-1.1819 , 1.0962)
Estimaciones	(-1.7025 , 0.9031)	(-1.1750 , 1.0955)
Valores reales	(0.5622 , -0.1422)	(-2.0627 , 0.0911)
Estimaciones	(0.5687 , -0.1415)	(-2.0545 , 0.0833)
Valores reales	(0.7387 , -0.6405)	(-1.4613 , 0.4721)
Estimaciones	(0.7431 , -0.6415)	(-1.4595 , 0.4637)
Valores reales	(-1.6012 , 0.6568)	(1.0365 , -0.8914)
Estimaciones	(-1.5975 , 0.6489)	(1.0405 , -0.8949)

Tabla 4.4 Comparación entre predicciones y valores reales

Para analizar la convergencia de las cadenas, en este caso se han generado tres cadenas por parámetro, en las figuras 4.21 hasta 4.32 se muestran los gráficos de las trazas de las mismas, apreciándose la estabilidad de las mismas.

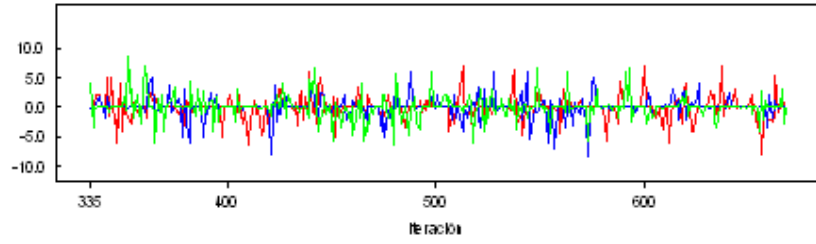


Figura 4.21 Traza de β_{10}

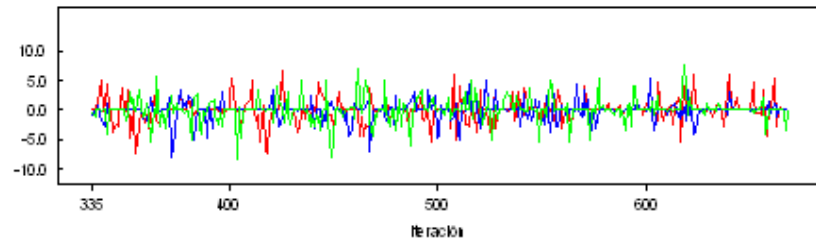


Figura 4.22 Traza de β_{20}

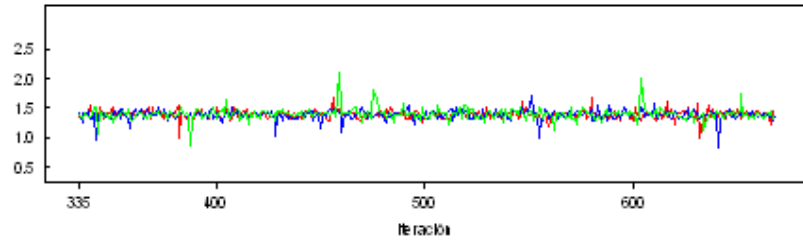


Figura 4.23 Traza de λ_{11}

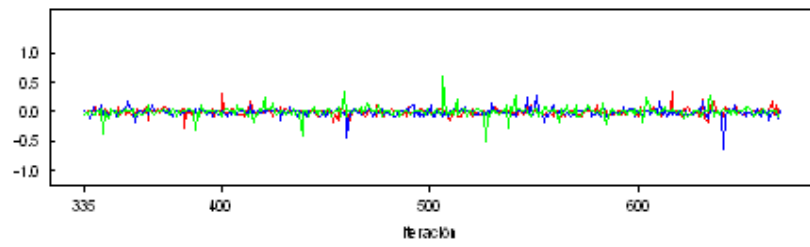


Figura 4.24 Traza de λ_{12}

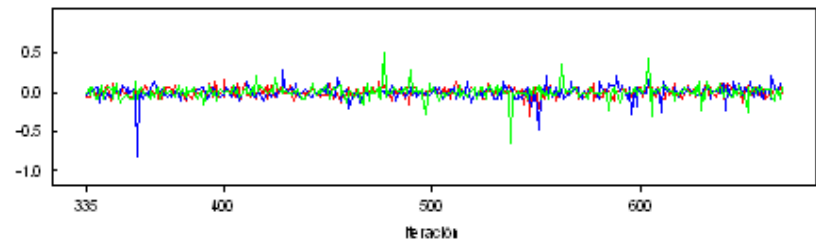


Figura 4.25 Traza de λ_{21}

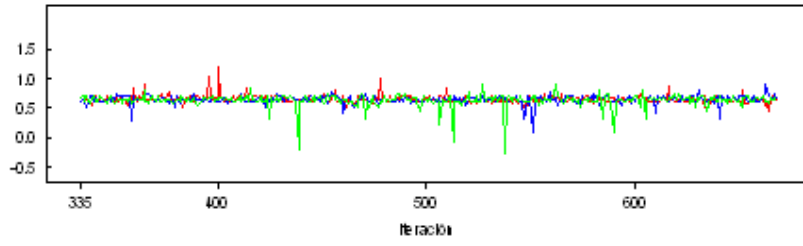


Figura 4.26 Traza de λ_{22}

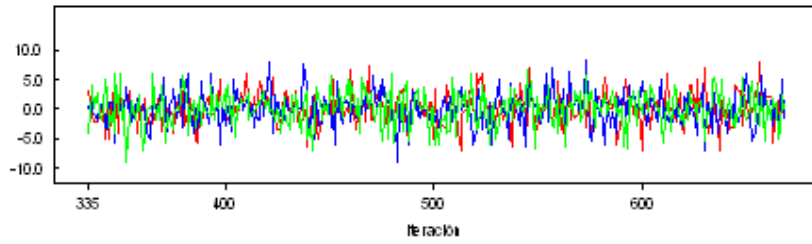


Figura 4.27 Traza de β_1

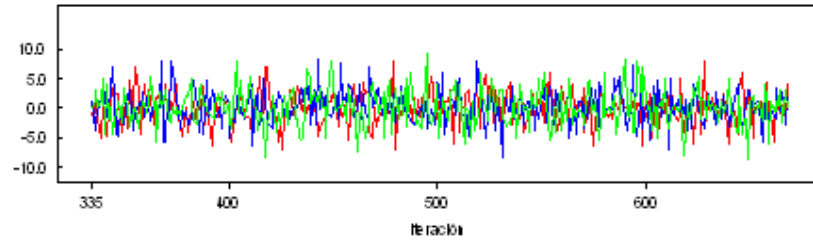


Figura 4.28 Traza de β_2

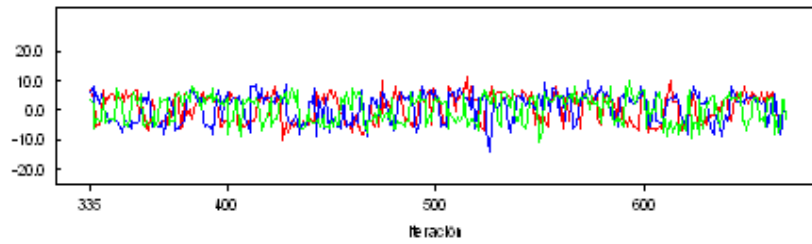


Figura 4.29 Traza de γ_0

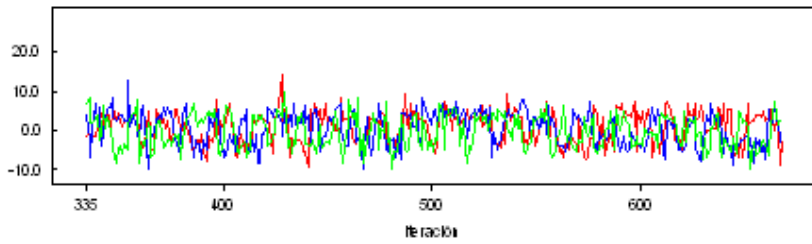


Figura 4.30 Traza de γ_1

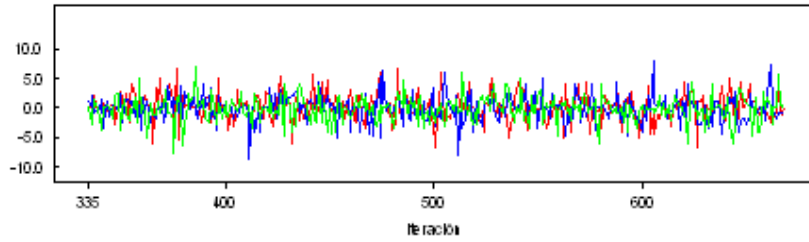


Figura 4.31 Traza de γ_2

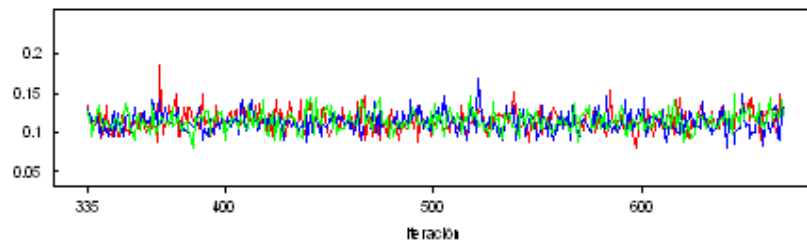


Fig. 4.32 Traza de σ^2

A continuación aplicamos el estadístico de Brooks, Gelman y Rubin, que se aplica en el caso de dos o más cadenas. En esencia, se trata de comparar la variabilidad dentro de las cadenas y la variabilidad entre cadenas, calculando el coeficiente *PSRF* (Potential Scale Reduciton Factor). Un coeficiente cercano a uno asegura que, en esencia, las cadenas se superponen, lo que indicaría convergencia. Aunque este estadístico se definió inicialmente para parámetros unidimensionales, Brooks y Gelman (1998), definieron una extensión al caso multidimensional para estudiar la convergencia simultánea de las cadenas de un conjunto de parámetros y éste será el que se aplique.

Por otro lado, también se utilizan los valores de los cuantiles del llamado *CSRF* (Corrected Scale Reduction Factor), corrección de PSRF, asumiendo que los parámetros se distribuyen según una normal. Un valor del cuantil 0.975 mayor que 1.2 se interpreta como evidencia de no convergencia.

En nuestro caso el valor del MPSRF es de 1.01, y la lista de los cuantiles del factor corregido para cada parámetro se muestra en la tabla 4.5. Todos ellos toman

un valor menor que 1.2, lo que la convergencia de las cadenas queda validada.

	cuantil 0.975 CSRF		cuantil 0.975 CSRF
β_{10}	1.02	β_1	1.00
β_{20}	1.00	β_2	1.01
λ_{11}	1.05	γ_0	1.01
λ_{12}	1.03	γ_1	1.1
λ_{21}	1.00	γ_2	1.0
λ_{22}	1.02	σ^2	1

Tabla 4.5 Cuantiles del CSRF para cada parámetro

Los resultados desarrollados en la sección 4.4 junto con su aplicación con los datos de la base ALL, se pueden encontrar publicados en Marín y Nieto (2008b).

Futuras líneas de investigación

A lo largo del desarrollo de esta tesis, han ido surgiendo nuevas ideas para abordar el problema de la superposición y emparejamientos de dos o más configuraciones y su extensión a situaciones más complejas.

Una primera continuación natural al trabajo realizado, es profundizar en los modelos no lineales de redes neuronales. Una posibilidad es extenderlos al caso de configuraciones en tres dimensiones ya que, por ejemplo, en el estudio de las moléculas de proteínas, la estructura tridimensional es muy importante. Otra opción es la inclusión de información adicional que se disponga de los puntos de las configuraciones a través de mediciones cuantitativas que permitan mejorar las superposiciones, es decir, trabajar con configuraciones con marcas.

Siguiendo con la idea de modelizar la relación entre dos configuraciones de puntos etiquetadas, otra posibilidad que consideramos interesante de investigar, es la modelización a través de splines de la transformación geométrica entre dos configuraciones o a través de procesos gaussianos como generalización de las redes neuronales. Esta idea podría utilizarse también en configuraciones no etiquetadas, añadiendo la dificultad de la estimación de los emparejamientos.

Por otra parte, el estudio del problema donde las configuraciones corresponden al mismo objeto en distintas situaciones hace pensar que el tratamiento debería ser distinto a cuando corresponden a objetos diferentes pues, en cierta medida, hay una *correlación* entre las configuraciones que se comparan. Equivaldría al clásico problema de comparar muestras independientes o muestras relacionadas. Como ejemplo,

pensemos en la imagen proveniente de un scáner de un paciente con cierta patología como una configuración de puntos, y el estudio de su evolución comparando las distintas configuraciones observadas en distintos instantes de tiempo.

Aunque las aplicaciones que se han presentado se han centrado en el ámbito de la Bioinformática, los modelos desarrollados en este trabajo podrían utilizarse en otros ámbitos de aplicación donde la información de la que se dispone pueda resumirse en configuraciones de puntos, etiquetadas o no. Pensamos que la búsqueda de situaciones reales y la colaboración con investigadores de estas áreas de aplicación, pueden dar lugar a trabajos futuros de gran interés.

Referencias

Andrieu, C., de Freitas, N. y Doucet, A. (2001). Robust Full Bayesian Learning for Radial Basis Networks. *Neural Computation*, **13(10)**, 2359-2407.

Benjamini, Y. y Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *JRSSB*, **57**: 289-300.

Borg, I. y Groenen, P. J. F. (2005). *Modern Multidimensional Scaling*. New York: Springer.

Broadbent, S. (1980). Simulating the key hunter. *Journal of the Royal Statistical Society. Series AJ*, **143**: 109-140.

Brooks, S. y Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulation. *Journal of Computational and Graphical Statistics*, **7**: 434-455.

Carlin B. P. y Louis, T. A. (2000). Empirical Bayes methods for data analysis. Chapman & Hall.

Carroll J.D. y Chang J.J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. *Psychometrika*, **35**: 283-319.

Celeux G., Forbes F., Robert C. P. y Titterton D. M. (2006). Deviance Information Criteria for Missing Data. *Bayesian Analysis*, **1**: 4, 651-674.

Cheng, B. y Titterington, D. M. (1994). Neural Networks: A review from a Statistical Perspective. *Statistical Science*, vol. 9, n° 1, 2-54.

Chiaretti, S., Li, X., Gentleman, R., Vitale A., Vignetti, M., Mandelli, F., Ritz, J. y Foa R. (2004). Gene expression profile of adult t-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. *Blood*, **103**(7): 2771-2778.

Cybenko, G. (1989). Approximation by superposition of sigmoidal functions. *Mathematics of Control Systems and Signals*. **2**: 303-314.

Devroye L. (1986). Non-uniform Random Variate Generation. New York. Springer

Dryden, I. L. , Hirst J. D. y Melville, J. L. (2007). Statistical Analysis of Unlabeled Point Sets: Comparing Molecules in Chemoinformatics. *Biometrics* **63**(1), 237-251.

Dryden, I. L . y Mardia, K. V (1998). Statistical Shape Analysis. Wiley.

Dryden, I. L . y Mardia, K. V. (1992). Size and shape analysis of landmark data. *Biometrika*, **79**: 57-68.

Evans, K. y Dryden, I.L. (2008). *Bayesian matching of unlabelled point sets using Procrustes and configuration models*.

Tech. Rep. <http://www.maths.nott.ac.uk/personal/ild/papers/evandryden.pdf>.

Galileo (1638). *Discorsi e dimostrazioni matematiche, in forma di due nuove scienze attenenti alla meccanica e i movimenti locali*. appresso gli Elsevirii; Opere VIII.

Gelman A., Carlin, J. B., Stern H. S. y Rubin D. B. (2004) Bayesian Data Analysis Second Edition. Chapman & Hall.

Gentleman, R., Carey, V. J., Huber W., Irizarry R. A. y Dudoit S. (Eds) (2005). Bioinformatics and Computational Biology Solutions Using R and Bioconductor.

New York: Springer.

Geweke, J. (1992). Evaluating the accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. *Econometrica*, **24**, 1317-1399.

Green, P. J. y Mardia, K. V. (2006). Bayesian alignment using hierarchical models, with applications in protein Bioinformatics. *Biometrika*, **93**(2), 235–254.

Green, P. J. y Sibson, R. (1977). Computing Dirichlet tessalations in the plane. *Computer Journal*, **21**: 168-173.

Grimson, E y Lozano-Pérez, T. (1987) Localizing overlapping parts by searching the interpretation tree. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**: 468-482.

Hsuan, F. (1979). Generating uniform polygonal random pairs. *Appl. Statist.* **28**: 170-172

Hurley, J. R. y Cattell, R. B. (1962). The Procrustes program: producing direct rotation to test a hypothesised factor structure. *Behavioural Science*, **7**: 258-262.

Karaman M. W., Houck, M.L., Chemnick L.G., Nagpal S., Chawannakul D., Sudano D., Pike B.L., Ho V.V., Ryder O.A. y Hacia J.G. (2003) Comparative Analysis of Gene-Expression Patterns in Human and African Great Ape Cultured Fibroblasts *Genoma Research*, **13**: 1619-1630.

Khatri, C. G. y Mardia, K. V. (1977). The von Mises-Fisher distribution in orientation statistics. *J. R. Statist. Soc.B* 39, 95-106.

Kendall, D. G. (1977). The diffusion of shape. *Advances in Applied Probability*, **9**: 428-430

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society*, **16**: 81-121.

Kendall, D. G. y Kendall, W. S. (1980). Alignments in two dimensional random sets of points. *Advances in Applied Probability*, **12**: 280-424.

Kent, J. T., Mardia, K. V. y Taylor, C. C. (2004). Matching problems for unlabeled configurations. *In LASR2004 Proceedings: Bioinformatics, Images and Wavelets*. R. G. Aykroyd, S. Barber, y K. V. Mardia (eds). 33-36. Universidad de Leeds.

Kosowsky J. y Yuille, A. (1994) The invisible hand algorithm: Solving the assignment problem with statistical physics. *Neural Networks*, **7**: 477-490.

Lewis, P. A. W. y Shedler, G. S. (1979). Simulation of non-homogeneous Poisson processes by thinning. *Naval Res. Log. Quart.* **26**: 403-413.

Liu, Y., Li L., y Wang Y. (2004). Free Form Shape Matching Using Deterministic Annealing and Softassign. *Proceedings of 17th International Conference on Pattern Recognition, 23rd - 26th August 2004*, Cambridge, UK, vol. II, pp. 128-131.

Mardia, K. V., Edward, R. y Puri, M. L. (1977). Analysis of Centrl Place Theory. *Bulletin of the International Statistical Institute*, **47**:93-110.

Mardia, K. V., y Jupp, P. E. (2000). *Directional Statistics*. Chichester: Wiley.

Mardia, K. V., Kent, J. T. y Bibby, J.M. (1979). *Multivariate Analysis*. San Diego. Academic Press.

Mardia, K. V. (1989). Shape analysis of triangles through directional techniques. *Journal of the Royal Statistical Society, Series B*, **51**: 449-458.

Marin, J. M. y Nieto, C. (2008a) Spatial Matching of Multiple Configurations of Points with a Bioinformatics Application. *Communications in Statistics. Theory and Methods*, **37**: 12, 1977-1995.

Marin, J. M. y Nieto, C. (2008b) Bayesian non-linear matching of pairwise mi-

croarray gene expressions. *Working Paper 08-25. Statistics and Econometrics Series 07*. Universidad Carlos III de Madrid.

McCulloch, W.S. y Pitts, W. (1942). A logical calculus of the ideas immanent in nervous activity forms. *Bulletin of Mathematical Biophysics*, **9**: 127-147.

Mosimann, J. E. (1970). Size allometry: Size and shape variables with characterizations of the lognormal and generalized gamma distributions. *Journal of the American Statistical Association*, 65:930-948.

O'Higgins, P. (1989). *A morphometric study of cranial shape in the Hominoidea*. PhD thesis, University of Leeds.

Raftery, A. E. y Lewis, W. M. (1992). One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science*, **7**: 493-497.

Rangarajan, A., Chui, H. y Bookstein, F. L. (1997). The Softassign Procrustes matching algorithm. In *Information Processing in Medical Imaging*, J. Duncan y G. Gindi (eds), 29-42. Berlin: Springer.

Richardson, S. (2002). Discussion of Spiegelhalter et al. *Journal of the Royal Statistical Society, Series B*, 631.

Rosenblatt, F. (1958). The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, **65**: 386-408.

Ross, S. M. (1997). *Simulation*. Academic Press. San Diego.

Small, C. G. (1988). Techniques of shape analysis on sets of points. *International Statistical Review*, **56**: 243-257.

Stoyan, D., Kendall, W.S. y Mecke, J. (1995) *Stochastic Geometry and its Applications*. 2nd Edition. Wiley, Chichester.

Stoyan, D. y Molchanov, I. S. (1997) Set-valued means of Random Particles. *Journal of Mathematical Imaging and Vision* **7**: 111-121.

Taylor, C. C., Mardia, K. V. y Kent, J. T. (2003) Matching unlabelled configurations using the EM algorithm. *Proceedings in Stochastic Geometgry, Biological Structure and Images*, 19-21. R. G. Aykroyd, K. V. Mardia y M. J. Langdon (eds), 19-21. Universidad de Leeds.

Thompson, D. A. (1917). *On growth and form*. Cambridge University Press, Cambridge.

Umeyama, S. (1993). Parameterized point pattern matching and its application to recognition of object families. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15(1)**: 136-144.

Walker, G. (1999). *Robust, non-parametric and automatic methods for matching spatial point patterns*. PhD tesis. Universidad de Leeds.